THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR .... Joseph Ford Belanger ....

TITLE OF THESIS .... Reading Skill as an Influence on Writing ....

.... Skill ....

DEGREE FOR WHICH THESIS WAS PRESENTED .... Doctor of Philosophy ....

YEAR THIS DEGREE GRANTED .... 1978 ....

THE UNIVERSITY OF ALBERTA

READING SKILL AS AN INFLUENCE ON WRITING SKILL

by

© JOSEPH FORD BELANGER

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF SECONDARY EDUCATION

EDMONTON, ALBERTA

FALL, 1978

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Reading Skill as an Influence on Writing Skill submitted by Joseph Ford Belanger in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

ABSTRACT

This study asked: Does a change in reading skill produce a consequent change in writing skill? Previous researchers attempting to examine this question were unable to effect significant changes on reading measures and therefore could not proceed further. In the current study, statistically significant changes on the reading measure were confirmed as part of a concurrent experiment before the analysis of the writing samples was undertaken.

Eight intact classes of grade-nine and -ten students (n = 194) were randomly designated experimental and control groups. The Davis Series 2 Reading Test was administered and writing samples on assigned topics were collected in November, just prior to the reading treatment, and again in February and May, allowing three test comparisons: pre-mid, mid-post, and pre-post. The writing sample was analyzed for overall quality, syntactic density, T-unit length, and fluency (the number of words in each composition). Using Diederich's three-rater method, rater reliabilities of .86 were obtained on the quality analysis. A corrected version of Golub's Syntactic Density Score (corrected to remove the mathematical anomaly in scoring the original scale) was used to calculate the syntactic density. In a subsidiary analysis, a subsample of the compositions was analyzed for four major sentence errors: run-on sentences, fragments, errors in subject-verb agreement, and errors in pronoun agreement.

Three-way analysis of covariance was used to test the statistical significance of the 36 main null hypotheses examined in the study. Two-way analysis of variance was used to test the effects of the treatment on high-, mid-, and low-ability students. In addition, correlations were calculated between the reading and four major writing measures and among the four writing measures.

Analysis of the reading measure showed that the experimental groups were superior to the control groups at the .002 level of confidence on the pre-mid comparison and at the .001 level of confidence on the pre-post comparison. No significant differences were observed on the mid-post comparison.

Analysis of the four major writing measures showed only random significant differences between the experimental and control groups. The control group showed significant gains ($p = .05$) on three of the thirty-six null hypotheses tested. Non-significant differences favored the experimental group on approximately one-half of the remaining null hypotheses. The subsidiary analysis of major errors was the only writing measure to show promise. Although only one of the nine comparisons reached significance ($p = .05$), both the mid-post and pre-post comparisons favored the experimental group. Two-way analysis of variance showed that the ability level of the subjects did not appreciably influence the results on either the reading or writing measures.

Correlations between the reading measure and both the

quality and fluency measures were generally significant ($p < .05$), but those between the reading measure and the syntactic density and T-unit measures were low. Correlations between the reading and quality measures were much higher for the girls ($r = .57$ to $.69$) than for the boys ($r = .33$ to $.37$), but correlations between the reading and other three writing measures were not appreciably related to the sex of the subject.

Correlations between the syntactic density and T-unit measures ranged from .90 to .94, and correlations between the quality and fluency measures ranged from .34 to .48. However, correlations between either the quality or fluency measures and either the syntactic density or T-unit measures were very slight and often negative.

The study produced no evidence to suggest a causative relationship between reading skill and writing skill. The absence of changes on the writing measures despite significant changes on the reading measure suggested that the view that such a relationship exists may be erroneous, with consequent alterations of existing theory.

ACKNOWLEDGEMENTS

Deepest gratitude is expressed to Dr. R.G. Martin, committee chairman, for inspiration and guidance throughout the project. I would also like to thank Mrs. E.B.H. Martin who administered the S.O.S. Reading Technique to the treatment groups in the study, and Dr. J.E. Oster who offered helpful suggestions on both the original proposal and the final drafts of the report.

To the many others who contributed to the study, a sincere thanks. Specifically, I would like to express appreciation to the following:

Dr. S. Hunka, T.C. Montgomerie, and J. Anderson of the Division of Research Services for assistance with designing and carrying out the statistical analyses.

C.E. Collins, P.S. Atwal, A.S. Tovey, and S.J. Barr, raters, check-coders and proof readers.

S. Shewchuk, Mrs. C. Walter, Mrs. M. Tesan, and Mrs. C. Sutton, teachers, and Mr. A. Yates and Mr. B. Day, principals with the Edmonton Public School Board.

M.K. Epp for writing many small computer programs which lessened the burden of the subsidiary analyses.

The Canada Council and the Department of Secondary Education for financial support.

My wife Jennifer who helped in countless ways and our daughter Christine who endured three years as a dissertation orphan.

Mrs. J. Theander who typed the manuscript.

TABLE OF CONTENTS

Chapter Page

Chapter Page

Chapter Page

Chapter Page

Chapter Page

Chapter | Page

Chapter | Page

LIST OF TABLES

Table | Page

Table Page

Table | Page

Table Page

Table | Page

Table Page

Table		Page

Table		Page

Table		Page

Table		Page

Table | Page

Table | Page

CHAPTER I

AN INTRODUCTION TO THE STUDY

THE PROBLEM

> Most of us have assumed the truth of the statement that the more effectively a student reads, the more effectively he writes . . . but among the thousands of recent articles on the reading problem there is a noticeable lack of studies on the relationship between reading and writing.
>
> Wykoff (1943, p. 250)

A. Background

Student literacy has become a topic of increasing public concern, with educators and businessmen alike decrying the average high-school and university graduate's alleged inability to write clear, well-organized, conventionally spelled and punctuated prose. As Rosenbaum (1973) suggests:

> Literacy, or the lack of literate skills, overshadows and outweighs every other problem and need sensed by educators and clients of the schools (p. v).

Both the proliferation of remedial writing programs at the university level and the results of a number of national tests and surveys support those who urge more effective measures for improving the quality of compositions written by students. Whether there is really a current "crisis" or merely a current concern, hardly anyone disputes the desirability, if not urgency, of better instruction in writing.

Over the past 20 years researchers have investigated a number of methods of teaching students to compose written

themes, and these methods, of course, are receiving renewed current attention in the present climate of concern about writing. Studies in writing practice and revision of themes, studies in pre-writing activities, and more recently studies in transformational grammar and sentence-combining practice have demonstrated that the quality of student compositions can be improved. None of these methods, however, can claim to offer a panacea for all the ills of student compositions: each offers a part of the solution when used in concert with other methods.

One potentially promising avenue of investigation which has not been explored in depth is the relationship between reading ability and skill in written composition. Eurich (1931) posed a question on this relationship which had remained unanswered almost a half century later: does an increase in reading ability produce a consequent increase in writing ability? A good deal of correlational data supports the hypothesis that the decoding skills of reading and the encoding skills of writing are closely related, and the educational literature abounds with speculation on the relationship between the two. However, the assumption that a student's skill in written composition can be enhanced by increasing his skill in reading has not been verified by empirical research. Two investigators attempting to examine the relationship between growth in reading ability and growth in writing skill (Eurich, 1931; Schneider, 1971) failed to effect a measurable change in reading skill. Therefore, the

2

question of consequent growth in writing ability remained unanswered.

The current experiment overcomes the difficulty of these former experiments by having assured differences between experimental and control groups in reading growth before the writing experiment was undertaken. These reading results were obtained in a concurrent experiment using the same experimental and control groups as those in the current experiment. In a study designed to explore the findings of two previous experiments of a smaller scale, (Martin 1975a, b, c), Martin and Belanger (1977) used the S.O.S. Reading Technique with a much larger experimental group to determine the technique's power to increase student reading abilities, in contrast with a control group which had not experienced the technique. Although data for the reading and writing experiments were collected simultaneously, it was not until the experimental group's reading change was assured that the writing experiment became a promising project.

The current experiment, then, asks: When an experimental group shows reading gains which are significantly greater than those shown by a control group at or beyond the .05 level of confidence, does the experimental group experience, as a consequence, significantly greater growth in writing ability than the control group does when this growth is measured by overall quality, syntactic density, T-unit length, and fluency?

B. Overview of Experimental Procedures

To examine the relationship between measured changes

in reading ability and measured changes in writing ability, four pairs of intact classes, two pairs on the grade-nine level and two pairs on the grade-ten level, each pair taught by one teacher, were randomly designated experimental and control. Following administration of pre-reading tests and collection of pre-writing samples in the form of in-class compositions, the experimental group was administered a reading treatment. Parallel reading measures and composition samples were taken three months following the treatment and again six months following the treatment. This design offered three possible test comparisons to examine the temporal relationship between the reading and writing scores: were the writing changes prompt (the first three months)? delayed (the second three months)? or spread evenly over the entire six months? The compositions gathered in the study were analyzed for four features, overall quality, syntactic density, T-unit length, and fluency, and, in addition, a subsample of the compositions was analyzed for major sentence errors. The differences between the compositions of the experimental and control groups were compared by the statistical procedure of analysis of covariance; in addition, to test the effect of the treatment on varying levels of ability, the experimental and control groups were divided into high-, medium-, and low-ability levels and the scores on the writing measures compared by means of two-way analysis of variance. Finally, correlations between the reading and the writing measures were computed.

4

C. Definition of Terms

Terms central to this experiment include:

1. Overall Quality. The average value, between zero and ten, assigned to a composition by at least two independent raters using the general impression rating method. These ratings were then compiled using procedures suggested by Diederich (1974; detailed and documented in Chapter III).

2. General Impression Rating Method. A method of grading compositions whereby the rater reads a composition only once rapidly and then assigns the paper a value. This method has been discussed by Wiseman (1949), Giffen (1972) and Diederich (1974).

3. T-unit Length. As defined by Hunt (1965), the researcher who developed the concept, the T-unit is "the shortest grammatically terminable unit into which a connected discourse can be segmented without leaving any fragments as residue" (p. 34). As a general rule, a T-unit is one main clause with all the subordinate clauses attached to it or embedded within it, independent of the punctuation that the writer has used. Consequently, most fragments are attached to a surrounding sentence while most run-on sentences are divided into their major clauses. The T-unit, then, is a measure of a student's ability to expand simple sentences, an index of his linguistic maturity, and not a measure of his ability to follow the conventions of punctuation (discussed and documented in Chapter III).

4. Syntactic Density. A tabulation of the structural features of discourse which includes both clause and sentence length factors (T-unit length, the average number of words per subordinate clause) and grammatical features (auxiliary verbs, possessives, prepositional phrases). Although the product of these tabulations converts to a "grade level" on a scale provided by the author, in the current study the raw score was used rather than the converted score. The Syntactic Density Score is discussed and documented in Chapter III; a modification of the scale, required because of a mathematical anomaly in the scoring procedures on the original scale, is presented in Appendix D.

5. Fluency. A measure of composition quality suggested by Harris (1962) and by R. Martin (1968), fluency is simply the number of words written by the student (discussed in Chapter III).

6. Major Sentence Errors. These are limited to four types of sentence errors found in written discourse (sentence fragments, run-together sentences, errors in subject-verb agreement, and errors in pronoun agreement). The absence of these errors was found by R. Martin (1968) to be a predictor of "good" writing on the grade-ten level.

7. Pre-, Mid-, and Posttests. In the current study tests were given three months apart allowing for the following comparisons:

a. the pre-mid comparison: Comparison of the pretest

with the test written three months following the treatment, the mid-test.

b. _the mid-post comparison_: Comparison of the test written three months following the treatment, the midtest, with the test written six months following the treatment, the posttest.

c. _the pre-post comparison_: Comparison of the pretest with the test written six months following the treatment, the posttest.

8. _Reading Skill_. Reading skill is defined as the ability to understand written material as measured by the _Davis Reading Test_, Series 2, Forms B, C, and D.

9. _Writing Skill_. Writing skill is defined as the ability to compose thoughts on paper as measured by overall quality, T-unit length, syntactic density, and fluency.

D. _Questions and Hypotheses_

This study hypothesized that groups of experimental students who had shown significantly greater gains on reading measures than control groups did would, as a consequence, show significantly greater gains on writing measures than the control groups did. To test this hypothesis, nine questions were asked to examine three grade groupings (grade-nine students separately, grade-ten students separately, and grade-nine and grade-ten students combined) over three time periods (the initial three months: the pre- and midtests; the second three months: the mid- and posttests; and the full six months: the

pre- and posttests). For each of the nine questions, the compositions of the experimental and control groups were compared on four writing measures: writing quality, syntactic density, T-unit length, and fluency.

A concurrent experiment demonstrated that the experimental groups showed significantly greater growth in reading ability than the control groups did. In light of the reading changes the current study examined the following nine questions:

1. Over the initial three months of the experiment is the writing of the grade-nine and -ten experimental groups significantly different from the writing of the control groups when the writing is judged on the following four measures:
   a. overall quality,
   b. syntactic density,
   c. T-unit length,
   d. fluency?
2. Over the initial three months of the experiment is the writing of the grade-nine experimental groups significantly different from the writing of the control groups when the writing is judged on the following four measures:
   a. overall quality,
   b. syntactic density,
   c. T-unit length,
   d. fluency?

3. Over the initial three months of the experiment is the writing of the grade-ten experimental groups significantly different from the writing of the control groups when the writing is judged on the following four measures:
   a. overall quality,
   b. syntactic density,
   c. T-unit length,
   d. fluency?
4. Over the second three months of the experiment is the writing of the grade-nine and -ten experimental groups significantly different from the writing of the control groups when the writing is judged on the following four measures:
   a. overall quality,
   b. syntactic density,
   c. T-unit length,
   d. fluency?
5. Over the second three months of the experiment is the writing of the grade-nine experimental groups significantly different from the writing of the control groups when the writing is judged on the following four measures:
   a. overall quality,
   b. syntactic density,
   c. T-unit length,
   d. fluency?

6. Over the second three months of the experiment is the writing of the grade-ten experimental groups significantly different from the writing of the control groups when the writing is judged on the following four measures:
   a. overall quality,
   b. syntactic density,
   c. T-unit length,
   d. fluency?
7. Over the full six months of the experiment is the writing of the grade-nine and -ten experimental groups significantly different from the writing of the control groups when the writing is judged on the following four measures:
   a. overall quality,
   b. syntactic density,
   c. T-unit length,
   d. fluency?
8. Over the full six months of the experiment is the writing of the grade-nine experimental groups significantly different from the writing of the control groups when the writing is judged on the following four measures:
   a. overall quality,
   b. syntactic density,
   c. T-unit length,
   d. fluency?

9. Over the full six months of the experiment is the writing of the grade-ten experimental groups significantly different from the writing of the control groups when the writing is judged on the following four measures:
   a. overall quality,
   b. syntactic density,
   c. T-unit length,
   d. fluency?

For purposes of statistical analysis, each of the nine questions outlined above was translated into a null hypothesis. The .05 level of confidence was considered statistically significant. These null hypotheses are given in Appendix B with a detailed presentation of the writing analysis.

E. Assumptions

The following assumptions underlie the current study:

1. That learning experiences in other classes during the course of the study were similar for the experimental and control groups. On the grade-nine level, since the students travelled as a class and since with the exception of "options" the experimental and control groups shared the same teachers for other subjects, the other subjects that the students studied can be expected to introduce very little bias to the data. On the grade-ten level, on the other hand, students had a variety of teachers for other subjects, the

English class being the only class in which all students involved in the experiment shared the same teachers. However, since the distribution in other classes appeared to be random, it is assumed that this factor did not introduce bias into the experiment.

2. That students would put forth their best effort without external rewards such as a grade on the paper. It is common research practice not to return the pre- and posttests to the students, partly because this might introduce bias to the data and partly because of the logistics of copying and returning the papers (Buxton, 1959; Mellon, 1969; Gajadharsingh, 1970; O'Hare, 1973). It was assumed that these conditions were similar for the experimental and control students and therefore would not introduce bias to the experiment.
3. That the specific writing topics developed from the general areas in Bell's (1971) assessment of student writing preferences were equally challenging for the experimental and control students.
4. That the six-month duration of the study was adequate time to bring about significant writing changes. Although researchers such as Gajadharsingh (1970) have effected marked writing changes with as little as six weeks of instruction, Burton (1973) suggests that six months is the minimum time

within which writing changes can be reasonably expected.

F. Limitations

Formal generalizability of the results must be qualified by the use of intact classes. The sample seems representative of, but not formally generalizable to, classes of ninth- and tenth-grade students in the Edmonton area.

Generalizability of the results is further limited to the "normal" Alberta school curriculum as taught in the grade-nine and -ten classes over the course of the current experiment. A school program with a concentrated course in writing or a school program with considerable time devoted to free reading (as was the case in the O'Hare (1973) experiment) might be expected to produce different results if either of these variables interacted with the reading improvement.

A further limitation is imposed on the generalizability of the findings by the specific reading treatment used in this study. The reading treatment is not considered likely to be effective with those who recently learned English as a second language or those whose initial reading training was by synthetic phonics (Martin and Martin, 1974).

G. Significance of the Study

The significance of this study lies both in the examination of a question basic to the teaching of English and in the investigation of a potentially efficient and

economical method of improving students' writing abilities.

The question of whether or not an improvement in reading ability brings a consequent, measurable improvement in writing ability, first examined in an experimental setting almost a half-century ago, had remained unanswered by empirical research up to the time of this study. Previous studies attempting to answer the question had either failed to produce a measurable increase in reading ability or failed to use adequate measures of writing ability. The current experiment overcomes both these difficulties: the experimental groups' superior gains in reading ability over the control groups were ascertained before analysis of the writing samples was undertaken; and the writing measures used in the study include both the subjective measure of overall quality and the objective measures of syntactic growth, growth in fluency, and reduction of major errors. Therefore, this study offers the opportunity to measure growth in written composition as a consequence of growth in reading ability. Either confirmation or rejection of the hypotheses would provide useful information. Finding that growth in writing is not simply a natural consequence of growth in reading would establish a bench-mark and suggest other avenues of research into the relationship between reading and writing. For example, this would suggest the questions: What factors in conjunction with reading growth produce growth in writing? Is writing instruction more effective when coupled with reading growth? What kinds of writing instruction? Does an

increase in reading ability when coupled with free reading produce more growth in writing ability than free reading alone does?[1]

From a practical standpoint, if the hypotheses of the study were confirmed and the quality of student writing were improved as a consequence of improved reading skill, this study would offer both an economical method of improving student writing (the reading treatment, discussed and documented in Appendix A is a 40-minute, one-time technique) and support for the theory that reading be taught as an integral part of the English program. Currently in Alberta high schools reading is offered as a course separate from language and literature. This study, however, did not purport to offer a panacea for all the ills of student compositions: it tested one method to be used in concert with many others in teaching students how to write compositions.

H. Organization of the Study

The basis, procedures, and results of this study are presented in six chapters.

1. Chapter I. An introduction to and an overview of the study. This chapter presents the problem, a summary of the background,

---

[1]The experimental design, of course, precludes a definitive statement that measurable change in reading skill does not produce a measurable change in writing skill. As will be noted in Chapter VI, failure to reject the null hypotheses can lead only to the conclusion that under the conditions outlined in the current experiment, measurable change in reading skill did not lead to measurable change in writing skill.

the questions and hypotheses, the assumptions, limitations, and a statement of the significance of the study.

2. Chapter II. A review of related research. This chapter briefly outlines research on interrelationships among the language arts and examines in detail both theories and empirical data on the relationship between reading and writing.

3. Chapter III. The experimental procedures. This chapter discusses the selection of the sample, the treatment, the reading and writing measures, the evaluation procedures, and the statistical methods used in the study.

4. Chapter IV. Preliminary analyses. This chapter outlines the results of a number of analyses that were carried out prior to the examination of the hypotheses. These analyses include the success of various randomization procedures, rater agreement and check-coding, and a summary of the reading results.

5. Chapter V. Summary of the findings. This chapter presents a summary of the writing results, a summary of the relationships between the reading and writing scores, and a summary of the correlations between the reading and writing scores. A detailed examination of the reading and writing scores, discussed in terms of the 36 null hypotheses tested in this study, is presented in Appendix B.

6. Chapter VI. Conclusions and implications for further research.

CHAPTER II

RELATED RESEARCH

Of the many investigations of written composition discussed in the standard reviews of research, few have examined the relationship between written composition and the other language arts: reading, speaking, and listening. However, investigations of the relationships between reading and speaking, reading and listening, and speaking and listening suggest that the language arts are closely related, that skill in one area is generally paralleled by skills in other areas, and that improvement of skill in one area will be accompanied by improvement of skill in other areas. In addition, descriptive and correlational studies of the relationships between reading and writing suggest that reading and writing are also closely related skills. On the other hand, while some investigators have reported success in teaching sentence and paragraph structure and in using sentence-combining exercises to improve reading, empirical studies attempting to improve writing through teaching reading have been generally unsuccessful, largely because of the investigators' lack of success in improving reading skill.

Greene (1941, 1950), Searles and Carlsen (1960), Meckel (1963), Braddock, Lloyd-Jones, and Schoer (1963), Braddock (1969), Sherwin (1969), and Blount (1973) have reviewed and discussed important research in the teaching of written

7

composition over the past century. These summaries report studies on traditional grammar and usage, writing frequency, curriculum, composition assignments, vocabulary, student characteristics, and teacher preparation. Research conducted subsequent to these summaries has introduced a number of new topics including prewriting activities, transformational-generative grammar and sentence combining, and descriptions of the composing process. Although this research has contributed a great deal to the understanding of teaching composition, it is only tangentially related to the current examination of reading and writing ability and will be included in the current review only when it applies directly to the current study (evaluating compositions or the effect of sentence-combining practice on reading skill, for example).

A notable omission in these reviews is the relationship between reading and writing skills. Although the profession showed considerable interest in the 1950's and 1960's in the interrelationships among the language arts, this interest is not reflected in research into the relationship between reading and writing skills. The only reviewer to devote a section to the interrelationships among the language arts, Meckel (1963), mentions only one study on the relationship between reading and writing - a correlational study by Diederich (1957) - and devotes a scant 150 words to the topic.

A. Interrelationships Among Reading, Listening, Speaking, and Grammar

Artley (1948), Hildreth (1954), Artley *et al.*, (1954),

Burrows et al., (1961), Meckel (1963), Russell and Fea (1963), and Harris (1969) have summarized research into the relationships among the language arts. In addition, the yearly "Summary of Investigations Relating to Reading" by Weintraub, et al., in Reading Research Quarterly includes recent studies on this topic. While many of the topics reviewed in these summaries (handwriting, spelling, letter writing, personality factors, academic performance, and bilingualism) are not of direct pertinence to the current study, other topics (the relationship between reading and grammar, reading and speaking, reading and listening, writing and speaking, writing and listening) point out relationships among language skills that lend at least peripheral support to the hypothesis that reading and writing are closely related skills and that enhancement of one skill might well have a direct effect on the other skill.[1]

Artley (1948) suggested a general rationale for expecting interrelationships among the language arts:

> Figuratively stated, speaking and reading comprise two sides of a square known as communication or language, the other two sides being writing and listening. Being inextricably associated, any limitation or facility in one is reflected to some degree in the others (p. 351).

Commenting on parallels in speaking and writing, Furness

---

[1]This section on the interrelationships among the language arts is not intended to be an exhaustive review of the research. Rather, it is intended to review major avenues of investigation which show how the language arts are interrelated and how altering one language skill may affect another. The section attempts to present a balanced view, outlining both studies which show and studies which fail to show strong relationships among the language arts. Section C examines in-depth the relationship between reading and writing.

(1960) suggests:

> What one learns in one form of language study transfers to others. The writer takes advantage of what he learns in speaking. The speaker profits by skills he acquires in the practice of writing. Each skill helps the other (p.265).

Buckingham (1940) contended that deficiencies in general language ability and not merely deficiencies in reading ability were the basic problems for the remedial reader. He suggested that when working with remedial readers "attention should be given to the whole task of learning symbolism."

Wilkinson (1969) and Ruddell (1965) suggested that oral training could provide the basis for learning to read. Wilkinson, examining reading-readiness activities, commented that:

> Oracy is the basis of literacy. There is little point in trying to teach children to read if they lack certain basic skills. We need remedial oracy, rather than remedial reading classes (p. 103).

Ruddell agrees. He suggests that:

> Reading comprehension is a function of the similarity of patterns of language structure in the reading materials to oral patterns of language structure used by children (p. 270).

Loban (1966) also noted the relationship between speaking and reading ability:

> Those who have the advantage of good oral proficiency (and who come from favored socioeconomic circumstances) accelerate their reading ability as their schooling progresses. Those whose social background offers a limited linguistic start in school fall further behind as reading instruction continues (p. 92).

Tovatt and Miller (1967), as part of an extensive investigation of the effect of an oral-aural-visual approach to writing at the grade-nine level, correlated standardized reading,

writing, and listening measures with each other and with student compositions. They reported that the STEP Reading and STEP Writing test correlated .91, the STEP Reading and STEP Listening test correlated .82, and the STEP Writing and STEP Listening test correlated .86. On the other hand, correlations between the STEP Writing, Listening, and Reading tests with rated compositions written by the subjects were considerably lower, with coefficients of .60, .44, and .62, respectively. The interesting point here is that the reading test correlates slightly higher with the written composition than the writing test does, suggesting that the STEP Reading test is a slightly better measure of the students' composition abilities than the STEP Writing test is. However, the lower correlations between the standardized tests and the written compositions might have been influenced by the relatively low interrater reliabilities of .69 reported for the compositions.

Not all would agree, however, that these relationships have pedagogical implications. Keltner (1957), for example, argues that

> . . . although there may be some relationships among the four skills of communication, by and large there is not sufficient correlation between skills in speaking and skills in writing to justify instructional programs which attempt to improve one skill by training in the other (p. 282).

1. Relationships Between Reading Ability and Understanding Grammar. Although researchers have found statistically significant correlations between the ability to read and

21

explicit and implicit knowledge of grammar, attempts to improve reading ability through teaching grammar have generally failed.[1] On the other hand, one promising related avenue of research is the use of sentence-combining exercises to improve reading skills.

a. <u>knowledge of grammar and reading skill</u>: Gibbons (1941), Robertson (1966), Stoodt (1970), Shackford (1976), Taylor (1976), McCaffrey (1976), Weinstein and Rabinovich (1971), and O'Donnell (1964), investigated the relationship between knowledge, conscious or applied, of grammar and sentence structure and skill in reading. Most investigators reported statistically significant correlations between the two skills; however, the findings were not unanimous.

O'Donnell (1964) reported correlations of .46 between grammatical awareness and reading comprehension and .44 between awareness of sentence structure and reading comprehension following an examination of first-year college students. In a descriptive study of grade-seven and grade-eight students, Shackford (1976) found that students' knowledge of grammatical structure and their reading ability correlated at .67.

---

[1]Nor has grammar been a successful vehicle for teaching writing. As Braddock, <u>et al</u>., (1963) noted, "the teaching of formal grammar has a negligible or, because it usually displaces some instruction and practice in actual composition, even a harmful effect on the improvement of writing" (pp. 37-8). Reported correlations between knowledge of grammar and writing ability have not been high. Researchers as early as Rice (1903) reported negligible correlations between knowledge of grammar and writing ability.

Robertson (1966), examining students in grades four, five, and six, found a significant relationship between a subject's understanding of connectives and his ability in reading, listening, and writing. On the fourth-grade level, Stoodt (1970) reported high correlations between a student's comprehension of conjunctions and his reading ability.

Gibbons (1941) used a "disarranged phrase test" to measure the ability of 25 third-grade students to see the relationships between parts of a sentence. She found the ability to reconstruct the sentences correlated highly with reading ability (.72 $\pm$ .07) and noted that when the group was divided into high-, mid-, and low-ability on reading skill there were significant differences between the high- and low-ability students on sentence reconstruction tests. Weinstein and Rabinovich (1971) used nonsense phrases with "very good" and "average" grade-four students to examine the students' implicit knowledge of grammar. Two of the four lists that each group was asked to memorize were syntactically structured while two lists were not. They found that the very good readers were able to memorize the structured lists more rapidly than the unstructured lists while the average readers had equal difficulty with both lists. They concluded, therefore, that the very good readers could make use of the grammatical structures inherent in the sentence while the average readers could not.

Taylor (1976) examined the relationship between primary students' awareness of the acceptability of selected

grammatical and semantic structures and their reading achievement. She found that at the grade-one level a student's ability to identify both grammatically and semantically correct sentences correlated highly with his reading ability, but on the grade-two level only the ability to identify grammatically correct sentences correlated with reading ability. McCaffrey (1976), on the other hand, examining students in grades two, three, and four, found a low and insignificant correlation between a student's understanding of syntactic differences in matched pairs of sentences and his reading ability. Hoyt (1906) found little correlation between knowledge of grammar and either the ability to interpret a poem or the ability to write well. Following his study of 200 ninth-grade students he concluded:

> There is about the same relationship existing between grammar and composition and grammar and interpretation as exists between any two totally different subjects, as grammar and geography (p. 500).

b. teaching grammar, sentence structure, and sentence combining to improve reading: Blumenfeld and Miller (1966) Rinne (1967), Whittaker (1974), Akers (1976), Crews (1971), and Dykes (1976) have reported generally unsuccessful attempts to improve reading skill by teaching various kinds of grammar. On the other hand, Combs (1977), Levine (1976), and Hughes (1975) reported that sentence-combining exercises improved reading skills.

Allen (1964) and Fagan (1971) suggested that teaching sentence structures and transformational grammar would help

students become better readers. Fagan based his thesis on research at grades four, five, and six which showed that students' reading comprehension was directly related to the number and types of sentence transformations used in a passage.

Over a one-year period, Crews (1971) taught a linguistic-based grammar patterned on Allen's sector analysis to grade-four students. Although the experimental groups showed writing gains superior to those of the control groups at the .001 level of confidence, there were no significant differences between the two groups on the reading measure. Small but statistically non-significant reading gains favored the control groups. Rinne (1967), using a five-week pattern-awareness-training program with 364 remedial high school readers, found that increased pattern awareness did not result in increased reading comprehension.

In a nine-week study on the grade-two level, Akers (1976) reported no significant differences in either silent or oral reading abilities for any of three treatments: a basal reading program, a basal reading program plus 15 minutes daily of free reading, and a basal program supplemented by 15 minutes daily of structural analysis. Dykes (1976) reported no significant differences in reading comprehension between students who had received transformational-generative grammar instruction and those who had received traditional grammar instruction: over the eight months of the experiment the only reading differences between the two groups were on a vocabulary measure; this

measure favored the experimental group at the .05 level of confidence.

Both Whittaker (1974) and Blumenfeld and Miller (1966) reported that the study of linguistic-based grammars resulted in reading improvement. Whittaker (1974), in a 16 week study of 200 college freshmen, showed that linguistically oriented instruction (using Paul Roberts' Patterns of English) was more effective in improving students' reading abilities than was a course in programmed reading instruction (using the SRA College Reading Program One). The experimental groups' posttest reading scores were significantly (p = .01) higher than those of the control groups. Blumenfeld and Miller (1966) taught a Chomsky-based grammar to remedial college students and reported that the experimental groups showed significant improvement in reading ability at the .05 level of confidence following 15 weeks' instruction.

Following suggestions by J. Martin (1968), Mellon (1969), and O'Hare (1973), several researchers have demonstrated that sentence-combining practice has a positive influence on reading ability. Replicating the O'Hare study, Combs (1977) found that one result of sentence-combining exercises was increased reading comprehension. Following eight weeks of instruction, experimental grade-seven classes showed gains in reading comprehension superior to those of the control classes at the .001 level of confidence. Reading rate, on the other hand, was not significantly affected by the treatment. Levine (1976) used 96 questions in sentence combining with 112

grade-three students and reported that experimental students showed gains on a standardized reading test significantly greater than those of the control students at the .01 level of confidence. However, she also noted that there were no significant differences between the two groups on a cloze test. Hughes (1975) found that grade-seven students made large gains in writing fluency following sentence-combining exercises but that only the low- and mid-ability readers showed significant gains in reading comprehension. Fisher (1973), on the other hand, found that sentence-combining exercises did not influence the reading ability of 47 fifth-, seventh-, and ninth-grade students. Although the writing quality of the experimental groups' compositions was superior to that of the control groups' compositions at the .001 level of confidence following the treatment, there were no significant differences in the reading comprehension of the treatment and non-treatment groups on either a cloze test or a standardized reading test.

2. Relationships Between Reading and Speech. Artley (1948), Dawson (1954), Parke (1961), Meckel (1963), Russell and Fea (1963), and Harris (1969) have summarized earlier studies, the majority correlational, on the relationships between reading and speech. Although some studies have reported high correlations between the two variables, the findings are by no means unanimous. On the other hand, later studies which attempted to teach speech and measure reading have reported significant changes in reading skill for only random

subgroups and on only random subskills.

Dawson (1954) reported studies by Rossignol (1948), Hildreth (1949), Hughes (1950), Dow and Papp (1943), and Davis (1938), the majority of which reported a positive relationship between reading and speech. Hughes found correlations of .67 between reading ability and the complexity of speech. On the first- and second-grade levels, Rossignol found that reading performance varied significantly with speech, while Davis, also examining primary students, reported small but statistically significant correlations between the speech ages and the reading ages of the subjects studied. On the other hand, Dow and Papp found no significant relationship between reading ability and speaking ability. Artley (1948) summarized nine earlier studies examining the relationships between silent reading impairment and speech defects (mostly stuttering), noting that eight of these studies found the number of speech defects considerably higher in impaired readers than in non-impaired readers, with ratios ranging up to ten to one.

Parke (1961) reported studies by Parke (1945), Martin (1955), and Winter (1957), all examining the first and second grades and all arriving at the same conclusion: the relationship between speaking and reading is virtually negligible. Martin concluded that "growth in each variable followed an individual developmental pattern and was unrelated to other variables." Strickland (1962) reported similar findings in the primary grades but noted that by

grade six there were "some" significant correlations between complexity of sentence patterns in children's oral language and reading achievement. On the other hand, Shire (1945; in Meckel, 1963) compared the speech of good and poor readers on the first-grade level and found that poor reading correlated with an absence of elaborated sentences, an abundance of short sentences, a lack of connectives, and a lack of vocabulary variety. Hopkins (1976), again using first-grade students, found statistically significant correlations between reading and three of five oral language measures: average utterance length ($r = .28$ to $.38$, depending on the reading measure; she used five reading measures), average T-unit length ($r = .25$ to $.37$), and Syntactic Density Score ($r = .24$ to $.29$). However, she reported statistically non-significant correlations between the reading scores and the number of T-units or the number of words in garbles. Rodgers, et al., (1974) reported similar findings. Dividing 81 grade-two students into high-, medium-, and low-ability on the basis of extensive oral tests in grade one they found reading differences between the high-, and low-oral ability groups significant beyond the .0001 level of confidence. Differences between the high- and mid-, and mid- and low-ability groups were significant at the .05 level of confidence. On the other hand, Calvert (1971) compared the oral language of reading achievers and underachievers in grades five and six, finding no significant differences in the oral language of the achievers and underachievers in

reading in average T-unit length, subordination ratio, or kernel structure ratio (his measure).

In a related study, Ruddell (1965) demonstrated that phrasing reading materials in patterns that resembled oral patterns improved reading comprehension on the grade-four level beyond the .01 level of confidence.

Attempts by Campbell (1973), Drumm (1976) and an Australian research group (1972) to improve reading through oral instruction did not yield consistent results. The Australian Study, "Improving Reading Through an Oral Language Program" (1972), taught an oral program to 108 grade-five students on the premise that reading competence was influenced by primary language competence. Following seven weeks of instruction (45 minutes daily) below-average readers showed gains significant at the .05 level of confidence, while average and above average readers showed statistically non-significant gains. Drumm (1976) used oral language-experience activities, synonym generation, and kernel-sentence expansion with experimental groups of grade-one students. Following five months (20 hours) of instruction, no statistically significant differences in reading ability were found, although the experimental groups showed "promise" on some of the reading measures. Campbell (1973) taught over 600 grade-three and -four students standard usage through oral instruction and exercises 20 minutes per day over 9 weeks. Neither the reading nor the standard usage in oral presentations improved, but scores on the two measures

correlated beyond the .05 level of confidence.

3. Relationships Between Reading and Listening. Dawson (1954), Parke (1961), Meckel (1963), Becker (1963), and Thompson (1967) have reviewed earlier studies on the relationship between reading and listening. Studies generally report high correlations between listening and reading skills, much higher correlations than were reported in the previous section on listening and speaking skills. This might be expected since listening and reading are both decoding skills. Devine (1976), summarizing studies from various grade levels, reported correlations between reading and listening ranging between .57 and .82. Johnson (1974), examining 203 students in grades two and three, reported correlations of .56 between reading and listening measures. Olson (1976) found low but statistically significant correlations ($p = .05$) between 100 grade-three students' understanding of synonyms and homophones in contextual oral language and their reading ability. However, attempts to improve reading skills through teaching oral skills have generally been unsuccessful.

Becker (1963) listed seven earlier studies which showed statistically significant correlations between reading and listening skills. On the university level, Selover and Porter (1937) reported correlation coefficients of .66; Larsen and Feder (1940), .82; Brown (1948), .35; and Nichols (1948), .46. Becker *et al.*, (1958) found correlations between reading and listening for superior college freshmen considerably lower ($r = .28$) than they were for the entire

freshman population (r = .44). Goldstein (1940), examining adults, reported correlation coefficients of .78 while Young (1936) reported correlation coefficients of .80 for grade-four, -five, and -six students. Thompson (1967) listed correlation coefficients for an additional six studies: Blewett (1949), .39; Heilman (1951), .66; Nichols and Keller (1953), .35; Biggs (1956), .52; and Abrams (1966), .44. Thompson cautions, however, that these correlations must be interpreted with care: "Part of this correlation is error in measurement and part of it probably is a result of unidentified factors common to both listening and reading" (p. 137).

Bond (1935), Levey (1975), and Wood (1975) compared the listening ability of good and poor readers. Bond (1935; in Dawson, 1954) found significant differences between good and poor readers on the second- and third-grade levels in auditory acuity, auditory discrimination, and auditory perception. Levey (1975) compared 60 grade-four achieving readers with 60 non-achieving readers and reported correlations between reading and listening significant at the .01 level of confidence for the achieving readers but low and statistically non-significant correlations for the non-achieving readers. Examining post-secondary technical school students, Wood (1975) found non-significant differences in the listening ability of high-, mid-, and low-ability readers when the rate of presentation was normal (175 words per minute) but significant differences favoring the high-ability readers when the rate of presentation was markedly

faster or markedly slower than normal.

Hall (1947), Larsen and Feder (1940), and Kirkham (1977) examined the differences between reading and listening comprehension. Hall (1947; in Dawson, 1954) found that neither listening nor reading was significantly more effective as a mode of assimilation of material by college students. Larsen and Feder (1940), in a detailed examination of the reading and listening comprehension of 150 university freshmen, reported correlations of .82 between the two skills, indicating that the skills are closely related, but noted that reading scores were six percent higher than listening scores. Further examination of their data showed that students of low ability were almost equal on the reading and listening measures, but that students of high ability gained much more from reading than from listening while low-ability students gained more from listening than from reading. They also noted that while easy material was comprehended as well in either mode, reading was markedly superior to listening when difficult material was involved. Studying 72 subjects in grades three, six, nine, and twelve, Kirkham (1977) arrived at similar conclusions: good readers at each grade level gained more from reading than from listening while average and below average readers gained more from listening.

Investigators attempting to teach listening skills to improve reading skills, however, have reported inconsistent results. Lewis (1976) taught listening skills to 60 experimental students in grades one, two, and three, and

reported that while experimental students at all three grade levels showed improvement significant at the .05 level of confidence on the listening measures, only the grade-one students showed a significant improvement on the reading measure over the control group. Denson (1973) taught 134 kindergarten students listening skills while an equal number of control students listened to stories on tapes. She reported no significant differences between the two groups on reading readiness tests but noted that the data might have been confounded by the similarity of the experimental and control treatments. Schulwitz (1973), on the other hand, reported that nine first-grade experimental classes were superior to control classes at the .05 level of confidence on a reading measure following six months' listening instruction. Weidner (1976) read stories to grade-four students using five different intensities of presentation. She reported that only the most intense treatment (ten minutes per day for three months) produced significant gains on reading-listening comprehension measures. In a one-semester experiment teaching 44 lessons in listening to five first-grade classes, Binzer (1974) reported no significant gains in reading for high-, medium-, or low-ability experimental students and noted that only the average and above-average students showed significant gains on the listening measure.

B. Relationships Between Oral and Writing Skills

Some theorists proposed that since speaking and writing are highly related skills, anything that enhances one will

enhance the other. Others suggest that these skills are not related closely enough to have any practical pedagogical implications. Although correlational data suggests a close relationship between speaking and writing, investigators attempting to enhance writing skills by teaching oral skills have not been able to demonstrate consistent improvement in writing skills. Two related avenues of research which appear to hold promise are oral sentence-combining practice and the talk-write pedagogy of Robert Zoellner (1969).

1. Theories. Gay (1977), Lundstein (1976), Blau (1968), Strickland (1960), and Horowitz and Berkowitz (1967) have noted theoretical relationships between oral and written communication and have suggested practical applications for these relationships in the classroom. Theorists, however, are not unanimous in these opinions and writers such as Keltner (1957), Becker (1963), and Thompson (1967) argue that the relationships between speaking and writing skills are minimal and that practical applicability in the classroom is therefore limited.

Based on experiments on acquisition and reproduction of memory using 56 freshman university psychology students, Horowitz and Berkowitz (1967) suggested that "listening seems (logically and empirically) more closely allied to speaking, and reading seems more closely allied to writing," but noted significant relationships among all four skills. Lundstein (1976) contends that written composition needs to be tied to oral language and that "Conversation and 'free talk' are the

basis for reporting, story telling and retelling, both original and from other sources." He urges teachers to exploit the links between written and oral communication. One such link is suggested by Gay (1977). Gay postulates five benefits from the teacher's reading aloud from "quality literary works" for at least 20 minutes per day. Contrasting the richness of such writers as Tolkien with the simplicity of the basal readers and noting that students are exposed to the style of the latter rather than that of the former, she contends that "It is unreasonable to expect children to learn to do something that they have never seen or heard." Gay suggests that reading aloud to students will enhance student writing by

1. increasing vocabulary,
2. improving ability to distinguish between subtle shades of meaning,
3. improving the sophistication and complexity of sentence structure,
4. providing a sense of structure and organization, and
5. providing a motive for writing.

Stewig (1975) suggests that teachers should read aloud to high-school as well as primary-school students "to introduce them to forms of literature that they might not ordinarily read." He also recommends a minimum of 20 minutes of oral reading per day.

Strickland (1960) commented that "if the ideas a child

expresses orally are meager, immature, and lacking in clarity, his writing will exhibit all of these problems in even greater measure." Blau (1968) concurs. Arguing for teaching oral argument and persuasion in the high school he contends that "a lack of proficiency in oral expression is the basis for a lack of writing proficiency."

Keltner (1957), Thompson (1967), and Becker (1963), on the other hand, argue that the relationships between speaking and writing skills are small and that they have no implications for teaching. Keltner summarizes:

> Thus, although there may be some relationship between communications skills in terms of the psychology of language behavior, the relationship between skill in speaking and skill in writing is very small, physiologically, psychologically, and socially (p. 281).

2. Correlational Data. Keltner (1957), Becker (1963), and Radcliffe (1969a, 1969b, 1972)* have summarized earlier correlational data on relationships between speaking and writing. Keltner cites studies by McCrery (1951), r = .35; Knower (1947), r = .19; and Bushnell (1930), r = .42 - all comparing oral presentations and compositions of college students and all statistically significant. Becker adds studies by Greene (1941) who found correlations of .40 between grades in a college composition course and the average scores on three oral presentations and by Becker, *et al.*, (1958) who reported correlations of .15 between speech and

---

*Gratitude is expressed to Terry Radcliffe whose unpublished papers and bibliographies acquainted me with a number of sources in the speech area that I was not previously aware of.

theme ratings of 176 superior college freshmen. Both Keltner and Becker argue that these correlations are low and that they illustrate differences rather than similarities between oral and written communications.

Recent studies by Gibson, et al., (1966) and Gatlin (1974), both examining college freshmen, also report statistically significant correlations between speaking and writing measures. Gatlin found correlations between oral and written maturity as measured by the subordination index significant at the .001 level of confidence and noted that there were no significant differences between the oral and written compositions when they were compared using Roberts' seven basic sentence patterns. Gibson, et al., compared freshman themes with five-minute extemporaneous oral presentations and found that the two correlated using Flesch's "reading eaze" (sic) score (.47) and vocabulary (.39), both statistically significant beyond the .01 level of confidence.

Heider and Heider (1940) produced earlier evidence on the relationship between speaking (hearing) and writing. Comparing sentence structure of deaf and hearing children, they found that the sentence patterns of deaf children were similar to those written by much younger hearing children.

Cooper and Odell (1976), on the other hand, found that professional writers did not consider "sound" an important factor in their own writing. The eight professional writers in the experiment made very few references to sound when discussing their own revisions of their work. However, when

Cooper and Odell made editorial changes in papers these writers had written and then asked the writers to comment on the changes, the writers referred to "sound" in their acceptance or rejection of the changes on 24 of the 75 changes. The changes, of course, were made to impair the "sound" of the sentences and consequently such comments might be expected.

3. Teaching Oral Skills to Improve Writing. Swain (1950), Miller and Ney (1968), and Wiggins (1968) attempted to improve writing skills through oral instruction. Although all three studies report writing gains by the experimental groups, only the Miller and Ney study reported statistically significant gains on the writing measure. Miller and Ney used oral sentence-combining practice with grade-four students and reported that following the treatment the experimental students wrote significantly (.001 level of confidence) more complex sentences and a significantly greater number of words within a half-hour time limit than the control students did. They noted that a delayed posttest showed even greater growth in writing for the experimental students. Wiggins (1968) used a series of 40 daily 30-minute oral lessons (including choral reading, oral discussions, and oral drills on sentence structure and punctuation) with experimental grade-four students. Following the treatment he reported small and statistically non-significant gains favoring the experimental groups on both a standardized test (Iowa Test of Basic Skills) and a composition. Swain (1950) reported that university

freshmen receiving public speaking training made statistically non-significant gains on measures of grammar, diction, capitalization, and sentence structure.

4. Talk-Write Strategies. One proposed method of integrating speaking and writing is having students compose compositions into tape recorders. Radcliffe (1969), Moore (1971), and Davis (1975) have subjected Zoellner's (1969) talk-write pedagogy to experimentation with inconclusive results. Related studies by Tovatt and Miller (1967), Zanotti (1970), and Gatlin (1974), however, lend support to pre-writing activities based on oral strategies.

Zoellner claimed that a talk-write composition course would incorporate three necessary principles of the writing process: it would allow the student to see the writing act in progress, it would make writing a social event, and it would provide an audience for the communication - three principles, he contends, that the traditional think-write pedagogies lack, but which are characteristic of natural human communication. Classroom procedures for this pedagogy have been outlined by Radcliffe (1969a) and Wixon and Stone (1977). The essential features of this strategy are a talker-writer who attempts to relate his ideas and a socio-reinforcer who gives the talker-writer encouragement. The conversation is recorded and provides a pool of ideas for the writer when it comes time for him to compose his thoughts on paper.

Research on talk-write, however, has been inconclusive. Radcliffe (1969b) used six volunteer high-school seniors

matched with six randomly chosen controls in a seven-week pilot project designed to test talk-write pedagogy. Following the treatment, one writing quality measure favored the experimental groups at the .05 level of confidence while another favored the control groups (not statistically significant). Radcliffe noted that his results were confounded by the method of choosing the populations, the small size of the sample, and the very low interrater reliabilities. Moore (1971) reported that a one-semester experiment using freshman university students and following Zoellner's theory resulted in slight and statistically non-significant gains favoring the control groups on vocabulary, writing fluency, and writing quality measures. Davis (1975), on the other hand, in an informal high-school classroom experiment reported that slight and statistically non-significant gains on composition posttests favored experimental students who had worked in pairs, composing compositions into tape recorders in a language laboratory.

Tovatt and Miller (1967) and Zanotti (1970) in related studies had students compose essays into tape recorders without the use of a socio-reinforcer. The Tovatt and Miller study reported statistically significant gains for the experimental groups on the STEP tests but not the writing assignments. Tovatt and Miller used an oral-aural-visual treatment with 72 grade-nine students in one experiment and 80 grade-nine students in a second experiment. They claimed that the "proficient writer will say and hear at the subvocal level what he is writing" and designed the experiment to have

students compose into tape recorders and to listen to what they had said. Following a full year's instruction, the experimental students showed statistically significant gains on the STEP Reading, Writing and Listening tests and on a test of English usage, but compositions written by the students showed only statistically non-significant differences favoring the experimental students. Perhaps the differences on the compositions could, in part at least, be attributed to the relatively low interrater reliabilities of .68 they reported. On the other hand, perhaps the Hawthorne effect accounted for much of the gains on the STEP tests. The classroom conditions for the experimental group described by Tovatt and Miller suggest that the experimental students enjoyed a great deal of freedom and that the teacher was only a friendly guide to writing. The conditions for the control group were not mentioned but appeared to be traditional. Zanotti (1970) used 120 grade-six students in a similar experiment but controlled for the Hawthorne effect by having the control students read their completed themes into the tape recorder while the experimental students actually composed themes on the tape recorder. Following the three-month experiment, the experimental students wrote longer compositions and judges rated the compositions of the experimental students higher, but neither of the comparisons reached statistical significance.

Gatlin (1974) offered related support. He compared the oral and written compositions of 92 freshman university

students. Half of the students made the oral presentation and then wrote the composition while the other half did this in reverse order. Gatlin noted that the students' compositions were more mature as measured by the subordination index if they talked before they wrote than if they wrote before they talked. This was significant at the .05 level of confidence. However, he did not report significant differences between the two orders of presentation on either Roberts' seven basic sentence patterns or the fluency (word count) measure.

C. Relationships Between Reading and Writing

Although a great deal of interest has been expressed in interrelationships among the language arts, few investigators have examined the relationships between reading and writing. As mentioned in Chapter I, 35 years ago Wykoff (1943) commented on the paucity of research examining traditional notions on the relationship between reading and writing skills. Wykoff cited only one such investigation (Eurich, 1931, discussed below).

Several theorists, however, have speculated on the relationship between reading and writing skills. Smith (1968), for example, suggested that reading and writing are clearly related skills which function through visual symbols and that "it might be hypothesized, therefore, that enhancing proficiency in reading contributes to proficiency in writing" (p. 5). In 1963, Braddock, *et al*., included the relationship between reading and writing under the heading *Unexplored Territory*: questions apparently "almost untouched by careful

research" (p. 52). Recently, Applebee (1977) noted that "We have progressed just far enough beyond their query to recognize that this may be an interesting avenue of investigation" (p. 535).

Research into the relationships between reading and writing skills has generally been premised on one of two rationales: the model rationale and the language-ability rationale. These rationales, of course, are not mutually exclusive. The more common rationale, imitation, suggests that learning to write, in common with learning all language skills, is a matter of imitation, conscious or unconscious, and, consequently, wide reading provides a model for the development of composition skills. This rationale has support from both a long history of traditional pedagogies and a limited amount of empirical research. Language ability, the less common rationale, assumes that reading and writing are both part of a much larger general facility with language and that improvement in one language skill will automatically lead to improvement in the other. Although a small number of empirical studies have attempted to investigate this relationship, most have foundered on the inability to alter the first variable, reading skill.

Previous research on reading and writing skills has examined both model and language-skill approaches, the former receiving the greater attention. Other evidence on the relationships between reading and writing is provided by correlational data, and investigations into teaching

beginning reading along with a number of related investigations and informal studies.

1. Research in Reading and Writing Skills. Eurich (1931), Mathews, Larsen, and Butler (1945), and Schneider (1971) attempted to improve composition skills on the college level by teaching reading. No definitive conclusions could be drawn from any of these studies. Neither Eurich nor Schneider was able to effect a measurable change in reading ability; Mathews, *et al*., reported significant gains in reading for the experimental groups but relied on subjective impressions of writing growth rather than standard measures. Recent studies by O'Donnell (1974), Miller (1974), D. Campbell (1976), and Matt (1977) have attempted to teach writing in combination with reading or to teach specific reading skills and measure writing growth, but again, no measurable differences in writing skills were reported. Those attempting to improve reading through teaching writing skills have also been generally unsuccessful. Although Obenchain (1971), Reed (1967), and Nagle (1972) reported that writing programs resulted in significantly improved reading skills, researchers such as Oehlkers (1972) and Calhoun (1971) who attempted to teach writing and measure reading did not report significant reading changes.

Eurich (1931) conducted the most comprehensive of the above investigations, four 12-week experiments during a two-year period. In the initial experiment, 83 experimental students were taught vocabulary and paragraph reading skills

while 87 matched control students studied the normal freshman composition course. The experimental groups showed significant gains over the control groups on only two of the eleven measures used in the experiment: two vocabulary measures. The experimental groups showed significant gains on the vocabulary measure when 25 of the 100 words on the test had been studied in class and very significant gains when all 25 of the words on the test had been studied in class. The experimental and control groups made equal gains on four measures (reading comprehension, vocabulary which had not been studied in class, an "objective" English examination, and final marks in all courses). The control groups made significantly greater gains than the experimental students did on rate of reading and made slight but not significant gains on mid-quarter English marks, final English marks, retention in reading, and on an essay examination. In experiments two to four, vocabulary and paragraph reading skills were taught separately for twelve-week periods. In these three experiments, Eurich found only one significant change from the results reported in the first experiment: the experimental groups lost their superiority on the vocabulary measure during the term vocabulary was not taught.

In a 15-week study of remedial students on the junior-college level, Schneider (1971) devoted 20 of the experimental groups' 75 hours of instruction to training in reading. The control groups studied the "normal" course. The experimental group and the control group showed

almost equal pre-post gains on the three measures used in the study: a standardized reading test, a writing mechanics test, and an essay test. Schneider, therefore, could draw no conclusions on the relationship between growth in reading ability and growth in writing ability.

Mathews, et al., (1945) discovered that 41 of the 460 students who earned a "D" in Rhetoric 1, freshman composition, had scored as high on scholastic aptitude tests as the average of the students who had earned a "B" in the course. However, the average "D" student's reading score was below the grade-eleven norm. Therefore, the following term the "D" students were taught written composition through reading exercises. No control groups were used and no standard measures were taken. However, the instructors' subjective impressions were that

> students did better reading than the usual special sections [sections designed for students who had earned "D's" in Rhetoric 1] and that . . . themes were at least as good as the average special section yields (p. 501).

In a follow-up experiment, three experimental groups were taught composition through reading exercises and three control groups were taught the "normal" course. Analysis of the pre- and posttests revealed that the experimental groups made much larger gains in reading rate and comprehension (ten centiles on both measures) than the control groups did, but that the groups made almost equal gains on an English mechanics test. Composition gains, again based on the instructors' subjective impressions, favored the control groups. However, since the experimental groups were select groups of brighter-than-

average students who had reading and writing deficiencies, and since the experimental groups showed substantial reading gains, if there were a direct relationship between reading and writing, one might expect writing improvement to be obvious even with the very subjective measures reported here. Paradoxically, the groups showing substantial gains in reading showed little growth in writing.

D. Campbell (1976) taught an integrated reading-and-writing course to two sections of freshman composition students for one quarter (12 weeks) while two control sections studied only writing skills. At the end of the experiment, slight and statistically non-significant differences on the writing measures favored the control groups while slight and statistically non-significant differences on the reading comprehension measure favored the experimental group.

Miller (1974) taught 62 remedial-composition students in two experimental sections reading skills integrated with short writing assignments and such practical vocational skills as the letter of application. Six sections used as controls (two of which were taught by the experimenter) studied descriptive and expository writing, the elements of rhetoric, and sentence usage problems. Following the one-semester experiment there were no significant differences between the experimental and control groups, either on a standardized reading test or the College Placement Test. On a vocabulary measure, the experimental students were superior

to the control students at the .05 level of confidence but on the writing-quality measure the control groups were superior to the experimental groups at the .01 level of confidence. It is clear from Miller's data, however, that the control groups received a great deal more instruction and practice in writing than the experimental groups did. This, combined with the fact that the reading instruction was apparently not effective, suggests why the experimental groups were less effective writers following the experiment.

Matt (1977) attempted to improve 40 high school senior's reading and writing skills through analysis of exposition and argument. Following the nine-week treatment he reported small and statistically non-significant gains in total writing scores favoring the experimental group. The experiment did not include a standardized reading test, but Matt reported a significant gain ($p = .05$) favoring the treatment group on a test he constructed to assess the students' ability to understand expository and argumentative prose, to infer the general purpose of the writer, and to infer how the writer accomplishes this purpose.

O'Donnell (1974) used two methods of teaching reading to 42 remedial college freshmen in an intensive four-week course. Neither the psycholinguistic nor the traditional reading method proved better at the end of the 14 one-hour sessions on any of the three measures: syntactic maturity, composition quality, or a standardized reading test.

Nor have studies attempting to teach writing skills

to improve reading skills been consistently successful. As reported above, Combs (1977), Levine (1976), and Miller and Ney (1968) reported that sentence-combining exercises led to improvement in reading skills. Investigators attempting to improve reading through teaching syntax and paragraph development, however, have reported inconsistent findings. Reed (1967), Obenchain (1971), Nagle (1972), and Weaver (1977) reported that writing instruction led to reading improvement, but Calhoun (1971) and Oehlkers (1973) reported that it did not.

Reed (1967) taught reading on the seventh-grade level through the study of syntax and paragraph structure. Following 15 weeks of instruction, the experimental students showed gains in comprehension superior to those of the control groups at the .01 level of confidence. Obenchain (1971) reported less definitive results. Seven experimental groups who were taught expository writing through the use of logical connectives wrote composition posttests superior to those of the control group at the .001 level of confidence. The reading results, however, were ambiguous. Four of the experimental groups, classes which had received the treatment throughout most of the year, showed reading gains significant at the .06 level of confidence, but three experimental classes, groups which had received the treatment during the first semester only, showed only slight differences in reading skill at the end of the second semester. Nagle (1972) used an experimental population of 371 eighth-grade students,

subjected the experimental students to a series of writing assignments, and reported that general reading was significantly improved (p = .05) in the treatment groups as measured by a standardized reading test. Weaver (1977) taught 31 third-grade students sentence organization through anagrams. Following the treatment (15 minutes per day, 3 days per week, over 3 months) the experimental students showed significantly (p = .02) greater gains on a cloze test than the control students did.

On the other hand, Calhoun (1971) taught effective writing techniques to 64 college freshmen over 10 weeks. Although the experimental group scored significantly higher (p = .01) than the control groups did on a posttest of "awareness of rhetorical techniques," there were no significant differences between the two groups on a post-written composition. Oehlkers (1972) used creative writing to teach 128 grade-one students "to encode oral language." Following the one-year experiment he found no significant differences on a reading measure between the creative writing groups and the control groups who were taught by a "language-experience" approach.

2. Wide Reading and Models. The use of models to teach writing can be traced to the Ancient Greeks. As Broudy and Palmer (1965) note, the Sophists felt that

> . . . the use of literary sources guaranteed an emotionally rich and evocative flow of language . . . a flexible command of the literature formed a common background of ideas, images, and feelings. . . . (p. 10).

In the English curriculum, the model theory goes back at

least to Benjamin Franklin, who "in his plans for a Philadelphia Academy (ca. 1750) had seen practical value in English literature as a model for writing . . ." (Applebee, 1974, p. 13). Not all writers, however, agree that the model theory is sound.

Theorists such as Frye (1962), Squire and Applebee (1968), Hughes and Duhamel (1967), Gleason (1965), and Perrin and Ebbitt (1972) suggest that reading enhances development of writing ability and general language competence. Frye (1962) contends that "it is no accident that the students who write best are usually those who read widely" (p. 41). Perrin and Ebbitt (1972) summarize this school of thought:

> Reading is probably the most valuable formative influence on a writer. Sometimes the influence is direct, as when there is a conscious, deliberate imitation of an admired author. More often, it is indirect, casual, cumulative. Just as we unconsciously pick up expressions and modes of expression from those we talk to, so we absorb rhythms, turns of phrase, and syntactic patterns from our reading. . . . If you read a good deal, you are likely to be more comfortable writing yourself; you will write more easily and confidently. And you will be a better judge of what you write (p. 26).

Others suggest that reading provides the student with materials to write about. Buxton (1966), for example, cites the British Columbia Curriculum Guide: "The best way to teach writing is to encourage the habit of reading so that the student may have a background of ideas . . ." (p. 104).

A number of educators have also speculated on the relationship between reading skill and the current problems with student composition. Biamonte, for example, states that ". . . no student is going to learn to write well if he

can't read. And with fewer and fewer of them reading, the situation is going to get much worse" (Adams, 1976). Veal suggests that "Writing is, after all, book talk. You learn book talk by reading" (Sheils, 1975).

Fenstermaker (1977) feels that many of the problems that students currently face in freshmen composition classes are the result of the students' lack of background in reading, lack of interest in reading, and inability to read critically. Fenstermaker suggests, however, that the lack of reading practice is most important:

> These students can read, but because they do not often read, they have little or no sensitivity to language, no understanding of nuance, no feel for structure. They lose the thread of any essay longer than two or three pages, regardless of its subject matter or style (p. 35).

Proponents of the model theory contend that select reading is required and that the student must be given models that illustrate certain principles. Direct imitation is generally stressed. Stern (1972) argues that "carefully guided reading of appropriate models can give the student of composition important insights into the logical and structural requirements of good writing." Hennings (1976) suggests that models can help students develop a sense of story patterns and sentence structure and also help in vocabulary development. She claims that models will help students "achieve a sense of the structure and sequence of clear communication." Squire and Applebee (1968) contend that the "judicious" use of models will help students achieve a better sense of direction and form in their writing.

D'Angelo (1973), Cooper (1973), Corbett (1971), and Winterowd (1970) are among at least 16 authors in the past decade to propose writing units based on literary models. D'Angelo's (1973) rationale is typical: the models provide a method for discovering alternative modes of expression and the student learns to write by imitating these. O'Dea (1965), on the other hand, feels that the student is overwhelmed by the difference between his own and the professional's writing and that this discourages the student. O'Dea argues that the student would benefit more by imitating the best writers in his class. Brown (1975), however, argues that models have no value in teaching composition and claims that there is no carryover from the study of rhetorical models. Arguing against the classical rationale he states:

> Yes, I know that Sophocles and Racine profited from what might be regarded as rhetorical models, so an exception might be allowed for students who want to write classical or neo-classical drama. Yes, I know Joyce received an elaborate training in rhetorical models, so another exception might be granted for the student writing the next Ulysses (p. 367).

Mills (1967), Ashida and Whipp (1963), and Andreach (1975) reported experiments based on the use of models to teach composition. Only one of these studies, however, the Mills study, reported definitive results. Ashida and Whipp argued that students learn the professional written dialect by counting the structures they find in professional writing and then trying to imitate these structures both in form and number. In an informal study with freshman composition

[illegible]

students, Ashida and Whipp compared this, the "slide-rule" method, with the "normal" composition course which used readings and discussions as activities to generate compositions. Their subjective impressions were that the experimental students improved more in composition during the semester than the control students did, but the researchers offered no standard measures of composition quality.

Mills (1967) used literary models with 45 grade-four and grade-five students two hours per week over 24 weeks. Control students spent the time with the "normal" school program. At the end of the experiment Mills reported that the writing sample of the two groups was "different" at the .001 level of confidence but that there were no significant differences between the two groups on the STEP Writing test.

Andreach (1975) used literary models with one class of grade-ten students while another grade-ten class studying the standard grade-ten English curriculum acted as a control. Using a rating scale he designed, Andreach rated the compositions himself. At the end of one semester, he reported that the experimental group showed gains greater than those of the control group at the .001 level of confidence, but he qualified the results by noting that:

> The improvement of the experimental group was attributed to the fact that both the experimental treatment and the rating scale concentrated on the single expository component of organization (p. 112).

3. <u>Correlational and Descriptive Studies</u>. Loban (1963, 1966, 1967), Evanechko, <u>et al</u>., (1974), Calhoun (1971), Fishco

(1966), D'Angelo (1977), Schonell (1942), Piexotto (1946), and Bippus (1977) reported statistically significant correlations between reading and writing skills. Lazdowski (1976) and M. Campbell (1976) found that reading ability could be predicted from a writing sample. Zeman (1969), Johnson (1976), Perron (1976), Thomas (1976), Harris (1975), and Kuntz (1975) showed that sentence types found in the compositions of high-ability readers differed from those found in the compositions of low-ability readers, but Baker (1954), Siedow (1973), and Fuller (1974) found no significant differences in the quality of compositions written by high- and low-ability readers. Researchers have also reported statistically significant correlations between writing skill and reading environmental factors. Donelson (1967), Barbig (1968), Maloney (1967), Monk (1958), Hyndman (1969), Woodward and Phillips (1967), and Lacampagne (1968) found high correlations between writing skill and the amount of reading that the subject did, the number of books found in the home, and the parents' reading habits.

Loban (1963, 1966, 1967), in his 13-year study, following students from kindergarten through grade twelve, found high correlations between reading and writing above grade two (he began to collect writing samples at the grade-three level). Commenting on the reading and writing abilities of grade-six students (1963) he noted that ". . . on every statistical measure one fact is extremely clear in the present study: those who read well also write well; those who read poorly also write poorly" (p. 75).

Reporting on the same students when they reached grade nine he stated that "Relationships between reading and writing become more pronounced as the years pass" (1966, p. 82). Evanechko, et al., (1974) compared the performance of 118 grade-six students on 8 reading sub-tests and 13 measures of syntactic maturity. They found that 76 percent of the correlations between language measures and reading measures were significant at or beyond the .05 level of confidence. One measure, Loban's "communication unit," correlated with the reading measures beyond the .00001 level of confidence. They concluded that their results suggested " . . . a very strong interaction between the receptive behavior of reading and the expressive behavior of writing" (p. 325). Fishco (1966), examining 95 seventh-grade students, found that reading comprehension and creative writing ability correlated beyond the .05 level of confidence. However, the correlation for the girls in the study was beyond the .01 level of confidence while the correlation for the boys did not reach the .05 level of confidence. Bippus (1977), examining students in grades four and six; Schonell (1942), examining students in grade-nine; and Piexotto (1946), Calhoun (1971), and Thomas (1976), examining college-level students, all reported statistically significant correlations between reading comprehension and writing achievement. The highest correlations ($r = .50$) were reported by Schonell and the lowest ($r = .13$) were reported by Thomas.

High correlations have also been reported between

reading ability and the kinds of sentence structures used in student compositions. Harris (1975) reported correlations between reading and written syntax tests on the grade-two level to be .70. On the grade-two and -three levels Zeman (1969) found that scores on a reading test correlated beyond the .01 level of confidence with the number of compound sentences found in the student's theme, but found no significant correlations between reading scores and Roberts' sentence patterns. Perron (1976), examining grade-five students, found statistically significant correlations between reading comprehension and both T-unit and clause length. Johnson (1976) reported that correlations between reading comprehension and both the number of words per T-unit and the length of main clauses correlated highly (r = .40) at grades three, four, and five, but noted a small and negative correlation between reading achievement and the Syntactic Density Score. Thomas (1976), examining college freshmen, reported small (r = .18) but statistically significant (.03 level of confidence) correlations between reading achievement and syntactic maturity. Based on his data Thomas concluded that "The student's ability to read is only negligibly related to his ability to write." Kuntz (1975) examined correlations between reading ability and the ability to make sentence transformations on the grade-seven level. She reported that correlations ranged from .68 to .80, all significant beyond the .001 level of confidence, and noted that the more difficult transformations correlated the

most highly with reading ability while the least difficult transformations correlated the least highly. Evans (1977), on the other hand, examining students in grade eight, twelve, and the final year of university, found that a cloze test correlated inversely with sentence-combining ability. The cloze test was of value in predicting the number of words per T-unit on a written composition only on the eighth-grade level. M. Campbell (1976) compared the scores of 40 university freshmen on a standardized reading test and an in-class composition. She concluded that "The better writers are indeed the better readers, and the poorer writers are in fact the poorer readers" but found that correctness in organization and mechanics was more highly related to superior reading skill than fluency of ideas was.

Diederich (1957) reported that good measures of reading ability were the most trustworthy indicators of writing ability. He found that the verbal score on the Scholastic Aptitude Test was more highly correlated with teachers' ratings of students' composition skills than was any other measure. Lazdowski (1976), working with samples from junior high, senior high, and college, found that the reading level of a student could be predicted within one grade level with a reliability of .88 from the student's grade on a written composition. On the college level, D. Campbell (1976) found that writing ability could be predicted from reading scores. D'Angelo (1977) on the other hand, examining grade-nine students, reported that reading

and writing ability correlated beyond the .01 level of confidence, but that listening comprehension and listening memory were more effective as predictors of reading ability than informative writing was.

On the other hand, investigators comparing the writing ability of good and poor readers have not found consistently strong relationships between the two skills. In Australia, P. Martin (1977) compared the reading and writing skills of six technical-school boys aged 12 to 14 and reported no consistent relationships between the two skills. Although one subject was good on both measures and another poor on both measures, the other four subjects were good on one measure and weak on another. Martin added that there was a general relationship between the two skills and concluded

> Despite the small numbers studied, the evidence suggests that reading and writing are intertwined but in ways that are not easily predictable. Indeed, it seems that much depends on the individual's perception of the purpose or usefulness of reading and writing, and the extent of his motivation (p. 52).

Baker (1954), comparing the reading skill of 100 remedial and 100 non-remedial students in freshman composition courses, noted only slight differences between the two groups. Fuller (1974) reported no significant differences in the written responses of good and poor readers on the junior college level. Siedow (1973), comparing reading ability and syntactic maturity in compositions of students in grades four, eight and twelve, found no significant correlations between the two measures.

A number of studies have reported a strong

relationship between reading skills and reading-environmental factors. Donelson (1967) selected the top 124 and the bottom 127 student writers at the grade-ten level from a population of 1821 students on the basis of a writing sample. He administered a 68-item questionnaire to his sample. Statistical analysis revealed that the following factors differentiated between the good and poor writers beyond the .01 level of confidence: number of books owned by the students, number of magazines in the home, the amount of reading done by the father (but, interestingly, not by the mother). He also found that the good writers "read more widely and more frequently" than the poor writers but reported that this data "was not amenable to statistics" (p. 40). Monk (1958) found that superior grade-seven writers were likely to be "children whose leisure-time reading was intensive, whose parents also did considerable reading, and whose homes were well-supplied with books." Maloney (1967), examining grade-nine students, Barbig (1968), examining students in grades nine to twelve, and Hyndman (1969), examining upper-average grade-ten students, arrived at similar conclusions. Woodward and Phillips (1967) found that the poor college freshman writer (one who received a "D" or an "E" in the first-semester writing course) had a lack of interest in reading and writing, a lack of reading materials in the home, and a sparsity of writing and reading experiences in high school when compared with good college freshman writers. Wyatt (1960), on the other hand, reported only one significant correlation between

extensive reading and skill in writing on the grade-six level: the amount of reading done and spelling unusual words. She found no significant differences between extensive and non-extensive readers when four compositions were judged on mechanics, usage, and sentence structure. Lacampagne (1968) compared the winners and the runners-up of the 1967-68 NCTE Achievement Awards with "average" students from the same schools as picked by the teachers. His 41-item questionnaire revealed only "some" correlation between extensive reading experiences and superior writing ability.

4. Writing Skill and Method of Teaching Beginning Reading. On the elementary-school level a large number of studies have examined the relationship between writing skill and the method of teaching beginning reading, the majority comparing students taught by the Initial Teaching Alphabet (i.t.a.) and those taught by Traditional Orthography (T.O.). Chall (1967) pointed out that in the earlier studies early writing practice was an important part of the i.t.a., but not the T.O., students' program and therefore one would expect differences favoring the i.t.a. students at the end of the first year. Later studies, however, notably Mazurkiewicz's (1975) 11-year study discussed below, overcame this criticism.

Downing (1967), Carner (1971), and Trost (1971) measured the written vocabularies of i.t.a.- and T.O.-trained students at the end of grade two and reported that the i.t.a.-trained students used more advanced words and a greater variety of words than their T.O.-trained counterparts.

Downing and Trost reported these findings significant beyond the .001 level of confidence.

Fyfe (1965) and Shapiro and Shapiro (1973) rated student compositions for quality in addition to the vocabulary measures. Fyfe reported that the i.t.a.-trained grade-three students wrote significantly better compositions than T.O.-trained students while Shapiro and Shapiro reported that the i.t.a.-trained students at the grade-two level wrote significantly better papers when the papers were rated on content scales and overall communication scales. Both investigations were reported significant beyond the .05 level of confidence.

On the other hand, Nikos (1970) reported no significant differences between the vocabularies of i.t.a.- and T.O.-trained students by the end of grade three. Auguste and Nalven (1969, 1972) reported that i.t.a.-trained students wrote significantly ($p = .01$) more creative compositions at the end of grade two but not at the end of grade four. However, apparently a different sample was used for the grade-four study. Students used in the latter study were matched on reading ability.

Sandel (1967, 1970) and Magnuson (1968) compared oral and written responses of i.t.a.- and T.O.-trained students. Sandel concluded that the i.t.a.-trained grade-three student's writing more closely approximated his oral language capacity than did that of the T.O.-trained student. Magnuson made similar observations at the grade-one level.

Mazurkiewicz (1967, 1971, 1973, and 1975), Ackerman (1969), Stewart (1969), Folta (1968), and Settles (1967) examined the sentence structures used by i.t.a.- and T.O.-trained students. The most comprehensive of these studies, Mazurkiewicz (1973), followed three groups of students from grade one through grade ten. Mazurkiewicz (1973) reported that by grade six the i.t.a. students wrote more words and longer T-units and used a wider variety of grammatical constructions than their T.O.-trained counterparts did, although none of these reached significance at the .05 level of confidence. When these students reached grade ten, Mazurkiewicz (1975) administered the 16 subtests of the SRA Assessment Survey Skills Profile. The i.t.a.-trained students scored higher than the T.O.-trained students on 13 of the 16 subtests, but only the subtest on spelling approached significance at the .05 level of confidence. Stewart (1969) reported that the i.t.a.-trained students wrote longer T-units than the T.O.-trained students, and Ackerman (1969) reported that i.t.a.-trained first-grade students wrote longer sentences, longer stories, and fewer mazes than T.O.-trained students. Folta (1968) found that the i.t.a.-trained grade-one students used fewer "and's" and "because's," made fewer failures per 100 sentence-combining transformations, and used more embeddings than T.O.-trained students did, although these differences were not statistically significant at the .05 level of confidence.

Smith (1968), in a study conducted with grade-one

students in five different geographical areas, concluded that students taught to read with synthetic phonics were superior in both reading and writing performance to students taught with a "meaning emphasis" (including "analytic phonics"). On both the word-meaning and paragraph-meaning sections of the Stanford Achievement Tests in reading, students trained by synthetic phonics were superior to students trained by analytic phonics beyond the .001 level of confidence. The synthetic-phonics trained students were also superior beyond the .001 level of confidence on two composition rating scales: 1) clarity and completeness of communication, and 2) spelling, length, original ideas, and use of rhetorical devices. Quinn (1977) compared the "basal" and "language-experience" approaches to teaching reading at the grade-one level. She found that compositions written by students trained by the language-experience approach were significantly ($p = .01$) better than those of the students trained by the basal approach on the following features: complete sentences, difficult words, and creativity.

5. Related Research and Informal Data. Studies by Heys (1962), DeVries (1970), and Glazer (1973) suggest that free reading and literature study can be used to improve writing, and informal data collected prior to the current study suggested that improved reading skill would be accompanied by improved writing skill. Elley, et al., (1976), however, reported that free reading appeared to have no effect on writing performance.

Heys' (1962) experimental groups, one class each from

grades nine through twelve, wrote one theme each week for one academic year while his control classes devoted the same amount of time to free reading. Over the school year, the reading group made gains double those of the writing group on both the STEP Writing Test and pre-post compositions judged by College Entrance Examination Board raters. Heys, however, did not report the statistical significance of these results. DeVries (1970) reported similar trends in a nine-week study of grade-five students, although the differences did not approach statistical significance at the .05 level of confidence. Glazer (1973) found that oral literature study helped grade-four and grade-six students to create significantly better written stories than those of students who had no planned literature study. She concluded that her results "lend credence to the theory that work in one area of language either directly aids growth in other areas of language or enhances general language facility."

On the other hand, Elley, et al., (1976) reported inconclusive results from a three-year New Zealand study. Following one year's instruction, control groups devoting 40 percent of their English periods to free reading and another 40 percent to literature study wrote slightly better compositions than groups studying transformational grammar and literature and groups studying traditional grammar and literature did. At the end of the third year, however, the groups studying transformational grammar wrote slightly better compositions than those of the other two groups

although none of these differences approached statistical significance.

As part of the initial design of the current study, informal data were collected from teachers who had used the S.O.S. Reading Technique[1] in their classrooms regarding the relationship between reading improvement and changes they had observed in the writing that their students produced. This informal data included both statements by these teachers and two small samples of papers written by students before and after the administration of the treatment.

6. Teacher Comments. Three Edmonton high school teachers who used the reading treatment in their classes during the 1974-75 school year made the following comments regarding the influence of the treatment on their students' written compositions:

a. In the better English 30 class, the writing has become more concise and easier to comprehend when marking. Therefore, the amount of marking time has decreased . . . one student, one of the better writers, claims she found research papers in other subjects easier to write.

b. One student was told by her social studies teacher that her last research paper was very clear compared to her earlier reports.

c. I am currently marking their Easter test papers and two particular boys are writing so well that had it not been a test I would be easily convinced it was not their own work. . . . Book reports are more academic.

d. Written work has greater clarity and better organization. The tendency to write an excellent answer to a question

---

[1]The S.O.S. Reading Technique, the treatment in the current study, is discussed and documented in Appendix A.

not asked has almost totally disappeared. Very poor writers were able to write more coherently and eliminate sentence fragments which usually abounded in their work.

e. In one session I corrected two sets of essays from my grade eleven class. The first set, which had been written before the technique was administered to the class, took almost twice as long to read and evaluate as the second set which had been written after most of the class had had the technique (in E. Martin, 1976).

7. Structural Analysis of Sample Papers. Informal analysis of two small samples of student compositions, one from grade-ten students and one from adults, reveals structural differences in the compositions written before and after the reading treatment was administered. Because the samples were very small and because no control groups were used, no statistical inferences can be drawn.

a. grade-ten compositions: In-class book reports were collected from three grade-ten students on September 18, 1975, in a school not used in the current experiment. Comparison of these papers, which were written before the reading treatment was administered, with book reports written by the same students December 9, 1975, three months following the administration of the treatment, reveals the following changes:

i. syntactic maturity: A calendar year's growth on Golub's Syntactic Density Score is considered 0.8 points; therefore, three months' growth ought to be 0.2 points. Comparison of the three pre- and three post-compositions reveals a growth of 0.313 points, or over 1.5 times the growth expected

over this three-month period.[1]

ii. <u>T-units</u>: Hunt's (1965) data suggest that between grades eight and twelve one year's growth in the average length of T-units is 0.73 words (p. 51). Comparison of the pre- and post-compositions of the grade-ten students reveals a growth of 1.11 words per T-unit, a growth of over 1.5 years over the three month period.

b. <u>adult compositions</u>: Informal in-class writing assignments were collected from post-secondary (non-university) students during a ten-week course in reading improvement. Comparison of the papers collected before the treatment was administered with those collected at the end of the course reveals a growth in average T-unit length of 2.67 words. The T-units written by superior adults in Hunt's sample were 20.3 words in length (p. 56). The adults in the above sample increased the average length of their T-units from 15.02 to 17.69 words over the ten-week period.

D. <u>Summary</u>

Most theorists agree that interrelationships among

---

[1]The scoring anomaly in the Syntactic Density Score (discussed in Appendix D) did not come to light in this initial analysis because ten T-units were selected from each paper. This follows a tradition established by Potter (1966), O'Hare (1973) and O'Donnell (1976) whereby a sample of ten T-units from each paper is analyzed rather than the entire paper. Since the number of T-units from each of the three essays was equal, the conversion score for the three essays seemed reasonable.

the language arts are strong. Students who speak well generally also read well and write well and are skilled listeners. The strongest evidence of interrelationships among the language arts is provided by Loban (1963, 1966, 1967) and Tovatt and Miller (1967). In his 13-year study of students from kindergarten to grade twelve, Loban noted not only that students who were skilled in one language-arts area were also skilled in the others, but also that the relationships became more pronounced as the years passed: each year the stronger students accelerated their progress while the weaker students fell further behind. Tovatt and Miller reported very high correlations among the STEP Reading, Writing, and Listening tests with coefficients ranging from .86 to .91, and high correlations between these tests and written compositions, with coefficients ranging from .44 to .62. Theorists such as Keltner (1957), Becker (1963), and Thompson (1967), on the other hand, argue that the correlations indicate differences rather than similarities among the skills and contend that the interrelationships have no practical application in the classroom.

1. Interrelationships Among Reading, Listening, and Speaking Skills. Although not all researchers have found statistically significant correlations between reading ability and understanding grammar, correlation coefficients as high as .67 have been reported between these two skills. On the other hand, those who have attempted to improve reading skill through teaching traditional grammars have been largely unsuccessful.

One promising avenue of research, however, appears to be the use of linguistic-based grammars and sentence-combining exercises to improve reading skill, with sentence-combining practice yielding the more consistent results.

Findings on the relationships between reading and speech are also inconclusive. While investigators have reported correlations as high as .67 between these two skills, nearly half of the studies reported a negligible relationship between reading and speaking, especially on the primary levels. Three attempts to improve reading skills through oral-language programs were largely unsuccessful: one study found improvement in reading for low-ability readers only while the other two studies reported no significant differences between the treatment and non-treatment groups.

Investigators have found that reading and listening, both decoding skills, correlate highly, the highest correlation coefficient reported being .82. Studies have also reported significant differences in the listening skills of good and poor readers. However, studies attempting to teach listening skills to improve reading skills have reported results inconsistent with each other. Two studies on the first-grade level showed that instruction in listening could significantly ($p = .05$) improve reading skill but two other studies on the first-grade level reported non-significant reading gains for the treatment groups.

2. Relationships Between Written and Oral Skills. Many

theorists argue that the relationships between oral and written communication suggest that the two skills should be taught together. Others, however, contend that the relationships between the two skills are not strong enough to have any pedagogical implications. Correlations reported between oral and writing skills, ranging from .15 to .47, are somewhat lower than those reported between either reading and speaking or reading and listening skills, but are all statistically significant beyond the .05 level of confidence. Using oral programs to improve writing has had mixed success. Miller and Ney (1968) reported that oral sentence-combining exercises significantly improved the written compositions of experimental students, but Wiggins (1968) and Swain (1950) found that oral programs produced statistically non-significant differences in compositions favoring the treatment groups. Experiments using oral pre-writing activities have also produced inconclusive results. Radcliffe (1969b), using Zoellner's Talk-Write pedagogy, reported that one composition measure favored the experimental group ($p = .05$) while another favored the control group (statistically non-significant). Two other experiments using this strategy reported statistically non-significant differences, one favoring the experimental and the other the control group. Two related studies using the tape recorder for pre-writing activities (but not the Talk-Write strategy) reported statistically non-significant gains favoring the experimental groups.

3. Relationships Between Reading and Writing. Over the past 50 years, 7 different investigators attempted to enhance skill in written composition by improving reading ability. Eurich (1931), Schneider (1971), D. Campbell (1976), Miller (1974), and O'Donnell (1974) failed to effect statistically significant changes in reading skill; consequently, the hypothesized attendant changes in writing skill could not be examined. Studies by Mathews, et al., (1945) and Matt (1977) were weakened by the absence of standard measures; however, neither found a statistically significant relationship between reading and writing skills.

Teaching writing skills to improve reading skills has also yielded inconclusive results. Reed (1967), Nagle (1972), and Weaver (1977) reported statistically significant changes ($p = .05$ or greater) in reading skill as a result of writing treatments, but Obenchain (1971) reported no significant ($p = .05$) gains in reading skill despite significant ($p = .001$) gains in composition skills resulting from a writing program. Calhoun (1971) and Oehlkers (1972) reported ambiguous results.

While the theory of using literary models goes back to the Ancient Greeks, not all educators agree that models have value in the classroom. Investigations using models to improve compositions have yielded inconclusive results. Following 24 weeks of instruction in the elementary school, Mills (1967) found significant differences in the written compositions of the experimental and control groups ($p = .001$) but found no

significant differences on the STEP Writing test. The lack of standard measures precluded definitive results from experiments by Ashida and Whipp (1963) and Andreach (1975).

Most investigators report statistically significant correlations between reading and writing skill. The correlation coefficients range from .13 to .80, depending on the measures involved. Fishco (1966) found that while the correlations between reading and writing skills were significant beyond the .01 level of confidence for the girls in his study, these correlations did not reach statistical significance at the .05 level of confidence for the boys. Investigators have also shown that scores on a reading measure can be used to predict writing skill and <u>vice versa</u>. Reading environmental factors such as the number of books found in the home and the amount of reading done by the father have also been shown to correlate significantly with reading skill, but researchers have not found statistically significant differences in the compositions of good and poor readers.

Comparisons of methods of teaching beginning reading also suggest a close relationship between reading and writing skills. Investigators have found that i.t.a.-trained students at the grade-one level have significantly larger written vocabularies ($p = .05$ and better) and write significantly better compositions ($p = .05$ and better) than T.O.-trained students. Smith (1968) also reported that synthetic-phonics trained students on the grade-one level were superior to

analytic-phonics-trained students on both standardized writing tests and written compositions beyond the .001 level of confidence. Beyond the primary level, however, the differences in reading training are not reflected in significant differences in composition skill. Mazurkiewicz (1975), for example, reported that although by grade ten i.t.a.-trained students scored higher on 13 of the 16 subtests of a standardized writing measure than T.O.-trained students did, the differences on only one subtest, spelling, approached statistical significance at the .05 level of confidence.

Two investigations have indicated that free reading is a more effective vehicle for improving composition skill than instruction in composition is. However, a third study of much longer duration failed to confirm these findings.

Informal data also suggested that the reading treatment used in the current study would lead to improved composition skill. Both statements from teachers who had used the treatment with their classes and a small number of compositions written before and after the treatment indicated that students who had received the reading treatment would show improvement in composition skill.

Evidence on the interrelationships among the language arts, then, generally supports the theory that reading, writing, listening, and speaking are highly interrelated skills. However, the hypothesis that instruction in one area will significantly alter skill in another has generally resisted confirmation. Research has also failed to establish

a causative relationship between reading and writing skills. Although some researchers have shown that writing programs significantly influence reading performance, researchers attempting to teach reading and measure writing have generally failed because of the inability to effect significant reading changes.

CHAPTER III

DESIGN AND PROCEDURES

The overall plan of the current study was to examine the relationship between measured changes in reading ability[1] and measured changes in writing ability over a six-month period at two grade levels. The experimental design called for two experimental and two control classes on both the grade-nine and -ten levels. Following the administration of the initial reading test and collection of initial writing samples, the experimental students were administered the reading treatment. Parallel reading tests were given and parallel writing samples were taken three months and six months following the administration of the treatment. Four measures were used to evaluate the compositions: raters evaluated a 37-percent sample of the compositions for overall quality, a 50-percent sample was analyzed using the Syntactic Density Score, and all the compositions gathered in the study were analyzed for T-unit length and fluency. In addition, a

---

[1]As noted in Chapter I, the current study was designed in conjunction with a reading experiment which was undertaken to examine the longitudinal effects of the S.O.S. Reading Technique, a replication on a larger scale of several previous studies (Martin, 1975a, b, c). The reading experiment was conducted and reported by Martin and Belanger (1977). However, since a number of students who had written all of the reading tests did not write one or more of the compositions, the sample sizes of the two studies differ. The reading results for the sample used in the current experiment are reported in Chapter IV.

37-percent sample was analyzed for major sentence errors. The statistical procedure of analysis of covariance was used to compare the scores of the experimental and control students on the four writing measures, and in addition, to test for effects at varying levels of ability, the experimental and control groups were divided into high-, medium-, and low-ability levels and the scores were compared using two-way analysis of variance. Finally, correlations between the reading and writing scores were computed.

A. The Sample

Four grade-nine and four grade-ten classes were chosen for this experiment. Five criteria for the selection of the classes in the sample were

1. that each teacher in the experiment teach one experimental and one control class;
2. that the experimental and control classes be deemed roughly comparable in ability by the classroom teacher;
3. that no concurrent school-wide reading or writing research be carried out in these classes during the time of the current experiment;
4. that the schools in the experiment not be considered atypical by the central office; and
5. that the schools be on a full-year rather than a semester schedule.

In consultation with the Director of Research, the Supervisor of English, and the Supervisor of Reading, one

senior and two junior high schools along with two alternate schools were chosen. Each of the four teachers, selected in consultation with the school principal, taught two classes at the same grade level which they considered roughly comparable. One class of each teacher was randomly designated experimental, the other control.

TABLE 1

The Experimental Population by Grade and Treatment

Grade/ Treatment	Original Population	Dropout/ Transfer	Missed Reading Test	Missed Writing Test*	Final Population
9 exp	56	2	2	2	50
10 exp	56	7	1	4	44
Total	112	9	3	6	94
9 con	57	2	0	2	53
10 con	55	4	1	3	47
Total	112	6	1	5	100

*Students who missed both a reading and writing test are counted only under "missed reading test."

As Table 1 shows, the original population of the experiment was 224 students, 112 experimental and 112 control. Over the six months of the experiment 30 students, 18 experimental and 12 control were excluded from the experiment. Fifteen of these students, 10 experimental and 6 control, either quit school or transferred to another school. Four students, 3 experimental and 1 control, missed at least one reading test while 11 students, 6 experimental and 5 control,

missed at least one writing test. The final experimental population, then, was 194 students, 94 experimental and 100 control. The large number of tests used in the study and the difficulties of June testing introduced a problem in the current study. Even with the examiner returning for make-up tests at least two days in each school (four days in the senior high school in June), 15 students missed one or more of the tests, 11 of these during the final June testing. Seven of these were grade-ten students, who, according to their teachers, were in school but simply skipped classes the day the examiner was there and one was a grade-nine student who merely signed his name to the paper. In addition, in June most students were observed to write only ten-to-twelve minutes rather than the average twenty minutes they wrote on the first two tests, a fact which is reflected in the lower fluency scores reported in Chapter V.

B. The Treatment

The Treatment used in this study was the S.O.S. Reading Technique, discussed and documented in Appendix A.

Immediately following the pretests in reading and writing, experimental classes were divided into groups of four or fewer for final groups if the number was not divisible by four. These groups were administered the treatment during consecutive one-hour periods over the next two school days. Administration of the treatment to the four experimental classes took approximately eight school days.

C. The Hawthorne Effect

Although research does not support the assumption that the Hawthorne effect operates in educational research, provision was made for the Hawthorne effect in the current experiment. In a two-year experiment Cook (1967) found no evidence of the Hawthorne effect in educational research, either in the 350 published studies he examined for evidence of the Hawthorne effect or in the 16-class experiment he conducted at the grade-four level attempting to induce the effect. Furthermore, Cook notes that in the original Roethlisberger and Dickson experiments the spurious gains were attributed to the extra attention paid to the experimental groups. In the current experiment, each experimental student received less than one hour's extra attention at the beginning of the experiment. At the three- and six-month testing periods each group received the same amount of attention. Nonetheless, provision was made in the experiment for the control groups to be given the opportunity to receive the treatment following the experiment, and the teachers involved in the study were so informed. Following the experiment only one student, a grade-nine girl, took advantage of this opportunity.

D. The Schedule

The initial reading and writing tests were administered November 25 to December 2, 1975, with make-up tests continuing until December 8. Administrations of the reading treatment began in a given school when testing was complete. The second testing period began February 24, 1976,

and, with make-up tests, continued until March 11. The final testing period began May 25 and continued with make-up tests until June 9, 1976. On the second and third tests the classes were tested in the same sequence in which they were tested on the first test, and each class received the reading and writing tests in the same sequence on the second and third tests as they did on the first test.

The second testing period was used to determine whether or not to proceed with the writing experiment. Since the experimental groups' reading gains three months following the administration of the treatment were superior to those of the control groups beyond the .005 level of confidence, the writing experiment depending on these differences appeared to be viable.

E. Tests and Measures

Three parallel forms of the Davis Reading Test, Series 2, were used as the reading measures in the study. In-class compositions written on assigned topics were the basis for the writing evaluation. Samples of these compositions were evaluated for overall quality and syntactic density and the full corpus was analyzed for T-unit length and fluency. In addition a sample of the compositions was analyzed for major sentence errors.

1. The Reading Measure. The Davis Reading Test, Series 2, was selected for the reading measure: 2D, the pretest; 2B, the midtest; and 2C the posttest. Coffman's review of the Davis

test (Buros, 1965) indicated that this was the best test with three or more forms available at the grade-nine and -ten levels. The Davis test had the additional advantage that it was used in the school district as a system-wide test for grade-nine students at the end of February. Therefore, the grade-nine students would not be subjected to two separate reading tests in February, which eliminated the possibility of one test confounding the results of the other. To assure consistency of administration, the schools allowed the investigator to administer the February test to the grade-nine students in the study.

The disadvantage of the Davis test, also noted by Coffman, is that it takes 45 to 50 minutes to administer, longer than the normal 40-minute periods in the schools in the study. Although this resulted in some inconvenience for all the grade-ten classes, all of the grade-nine classes had regularly-scheduled "double periods" and thus the test could be accommodated within the allotted class time.

2. The Writing Measure. One 30-minute in-class composition was written on assigned topics at each test period.

a. the number of writing samples at each test period: Research practice offers inconsistent information on the number of compositions required from each student for optimal evaluation of growth in composition. Researchers such as Mellon (1969), O'Hare (1973), and Gajadharsingh(1970) have suggested that more than one composition per testing period

is required. Based on Kincaid's (1953) findings, Mellon collected nine pre- and nine post-compositions while O'Hare collected five at each test period and Gajadharsingh collected three. However, examination of Kincaid's findings reveals that one must distinguish between "individual" and "group" performance. Kincaid found that

> a single paper written by a student on a given topic at a particular time cannot be considered as a valid basis for evaluating his achievement in a writing course at any time . . . ." (p. 86).

However, when he examined the performance of a group (n = 20) rather than that of an individual, he found that "there were no significant effects on the quality of writing as a result of variations in the day to day efficiency . . . or examination pressure" (p. 88). He concluded, therefore, that

> An evaluation of the overall, or average, group improvement resulting from a writing course may be obtained from a single pre-test theme and a single post-test theme (p. 95).

Diederich (1974) also suggests that one theme is adequate for the measurement of group performance. Although he feels that two short essays are required to establish an individual's writing ability, he recommends a single theme for comparison of class-sized groups (p. 13). Therefore, since the above studies were comparing groups of students and not individual students, the number of writing samples seems excessive, and in the case of Mellon's nine themes, the writing practice given by the tests might have had a confounding influence on the data.

The size of the experimental population and the

disruption of the school programs were two other considerations used in determining the number of writing samples to be collected. The current study had a population of 194 subjects. If the small populations studied by Kincaid (n = 20) showed statistically non-significant daily fluctuations in composition quality, then the much larger sample size of the current study (experimental n = 94; control n = 100) should be expected to show very little daily fluctuation. The six class periods, three for the reading tests and three for the writing tests, represent about three percent of the students' English class periods for the year. On the other hand, three reading tests and three writing samples at each of the three testing periods would require 18 classes or 10 percent of the year's English classes. The 9 writing samples which Mellon collected, if matched with reading tests, would take 30 percent of the English class periods for that year.

b. selection of composition topics: Writing topics assigned in this study were adapted from Bell's (1971) study of the writing preferences of senior high school students in the Edmonton area. Each of the three topics was assigned to one-third of the students at each testing period to avoid any systematic bias in topic presentation (Mellon, 1969; O'Hare, 1973).

Bell discovered that grade-ten students given a free choice of topics to write on chose "school," "society," and "home" most often as topics for their essays. These three

areas were used to develop the following topics:

"The Perfect (or Ideal) School"

"The Perfect (or Ideal) Family or Home"

"The Perfect (or Ideal) Country or Society"

Students were instructed to write on these topics in any way that they chose. It was suggested that they could choose real or imaginary situations and use real or imaginary details. A sample instruction sheet is given in Appendix O.

Topics were assigned in the combinations shown in Table 2.

TABLE 2

Order of Topic Presentation

Topic	First Test	Second Test	Third Test
School (J)	J	K	L
	J	L	K
Home (K)	K	J	L
	K	L	J
Society (L)	L	J	K
	L	K	J

Table 2 illustrates the six possible orders of topic presentation. For example, a student receiving topic "J" for the first test might receive topic "K" or topic "L" for the second test, but his third test topic was determined by the previous two. All possible orders of presentation were used in numbers equal except for the unequal division of the sample and the effect of dropouts.

c. <u>distribution of the topics</u>: One ditto master was made for each of the six possible topic combinations (JKL, JLK, KJL, KLJ, LJK, LKJ) with the code numbers typed at the top of the page. For the initial writing sample six copies of each topic combination were randomly intermingled for each class, a total of 36 sheets per class. These were then randomly distributed to the classes. For the second and third writing samples students were assigned topics according to the code on the top of the initial sheet. Students' names were written on the ditto sheets prior to the examination and the students were simply handed the topics with their names on them for the second and third samples.

d. <u>adjustment for topic difficulty</u>: When the compositions gathered in the study had been analyzed and the scores tabulated, a separate tabulation was made for each of the three topics: home, school, and society. A separate class list was made for each of the eight classes and each of the four writing measures and a color coded entry (Topic J was coded green; Topic K, red; and Topic L, blue) was made for each score that each student earned at each test period. The average score received on each topic was computed for the grade-nine level and for the grade-ten level. Computations were checked for accuracy by combining the scores on the three topics on each class list and checking these against computations made for the average score at each test period. A breakdown of the scores by topic, grade, treatment, and test period is presented in Appendix E. The computations

and adjustment factors are presented in Chapter IV.

In addition, to assess the influence of the adjustments for topic difficulty on the statistical computations, the "raw" or unadjusted scores of the experimental and control groups were compared for a sample of 8 of the 36 hypotheses tested in the experiment. Comparison of the results for the adjusted and unadjusted analyses are presented in Chapter IV.

3. Analysis of the Writing Sample. Compositions collected in the study were, as noted above, subjected to four separate analyses: overall quality, judged by experienced raters; syntactic density, following Golub's (1973) Syntactic Density Score; T-unit length developed by Hunt (1965); and writing fluency, suggested by Harris (1962) and R. Martin (1968). A 37-percent random sample of the papers was selected for the quality analysis and an additional 13 percent (a total of 50 percent) for the syntactic density measure. The entire sample of papers collected in the study was analyzed for T-unit length and fluency.

Because of the large number of compositions collected in the study, 582 in all excluding dropouts, time-consuming analyses such as the qualitative analysis and the syntactic density analysis would have proved difficult to carry out. The raters in the study would have required approximately 12 two-hour sessions rather than the 4 two-hour sessions they spent. Furthermore, the syntactic density score requires approximately one hour per thousand words to score. On the

other hand, for purposes of statistical analysis, sample sizes of at least 30 are considered desirable (Ferguson, 1971) in both the experimental and control groups. Therefore, samples were selected for the quality and syntactic density measures. The T-unit and fluency measures, on the other hand, were not time-consuming measures to score and so the full corpus was analyzed on these measures, as this would offer a check on the validity of the samples chosen. Thus, if the quality and syntactic density measures produced results varying widely from those produced by the T-unit and fluency measures, the sample and the measures could be examined to see where the differences lay.

a. sampling procedures: For the quality evaluation the compositions of nine students were randomly selected from each experimental and each control class, a sample size of approximately 37 percent of the subjects, a total of 108 experimental and 108 control compositions out of 582. The papers of an additional three students from each experimental and each control class were added for the syntactic density analysis, bringing the total number of compositions analyzed to 288, approximately 50 percent of the compositions gathered in the study.

Alphabetical class lists for each of the eight classes in the study were made and each class list was numbered from 1 to x. A table of 12 random numbers for each class list was generated by the HP 9830 computer at the Northern Alberta Institute of Technology. The first nine numbers in each

table were used for the quality analysis and the final three numbers were added for the syntactic density analysis.

To indicate the success of the randomization procedures, the initial reading scores, the three-month reading gains scores, and the six-month reading gains scores of the students selected for both the quality and the syntactic density random samples were compared with the scores of these in the remainder of the classes. The results of these analyses are presented in Chapter IV.

b. <u>the quality evaluation</u>: Three experienced raters[1] were asked to evaluate the 216 compositions selected for the quality analysis using a method suggested by Diederich (1974) and under conditions outlined by Braddock, <u>et al</u>. (1963).

i. <u>the environment</u>: The raters evaluated compositions from 9:30 to 11:30 a.m. on four consecutive days. To control for rater fatigue, a ten-minute coffee break was held one hour after each session started. Following the suggestion of Braddock, <u>et al</u>. (1963) the investigator was in the room with the raters to answer questions.

ii. <u>training</u>: Prior to the rating sessions the

---

[1]Raters A and B each had backgrounds as technical writers and editors and in addition both had taught writing for nine years in a technical school. Both were Certified Engineering Technicians in technical writing. Rater C had taught 24 years (the last nine in a technical school) and held B. Ed., M.A. and M. Ed. degrees.

three raters in the study spent two sessions reading and discussing papers gathered in the study but not included in the sample to be rated. These compositions, written by both grade-nine and grade-ten students, were written by students who had missed one or more tests and therefore could not be included in the study. The raters discussed these papers in terms of ideas, organization, style, vocabulary, and sentence structure, following suggestions outlined by O'Hare (1973), and used these papers to decide on a grading system. On the basis of the sample papers the raters, with the concurrence of the investigator, decided to use a modified percentage system: each paper would receive a grade of from zero to ten. In addition to the training sessions where 15 papers were read and discussed, each rating session was begun by rating three practice papers, grade-ten practice papers on the two days that grade-ten papers were rated and grade-nine papers on the days that the grade-nine papers were rated.

iii. procedures: Following procedures outlined by Diederich (1974, pp. 14-15) raters A and B were asked to rate the 216 compositions selected for the quality analysis by the overall impression method. If the ratings were not more than one point (out of eleven) apart, the paper's rating was the average of the ratings given by the two raters. If the ratings were more than one point apart, the paper was rated by Rater C and his rating was averaged with that of the rater most closely agreeing with him. If the grades of Raters A and B were equally distant from Rater C, the rating of

Rater A and B nearest the sample mean was discarded. Diederich suggests that this procedure is necessary because combining or averaging grades "pushes everyone toward the middle" (p. 20). Following Buxton's (1959) suggestion, raters were asked to compare ratings of the twenty-fourth and twenty-fifth papers rated each day to check that they were using similar criteria but not to alter the grade of any given paper. Agreement among the raters is discussed in Chapter IV. The re-rating of a subsample of the papers and the rationale behind re-rating them is presented in Appendix C.

c. syntactic density: A 50-percent sample of the compositions collected in the study was analyzed using a revised version of Golub's (1973) Syntactic Density Score, a measure which has the T-unit length (Hunt, 1965) as a basic component but goes beyond clause and sentence length factors and includes such items as the number of modals, the number of "be" and "have" auxiliaries, the number of prepositional phrases, the number of possessives, the number of adverbs of time and the number of gerunds, participles, and absolute phrases found in a writing sample. In Golub's scale, each grammatical feature is tabulated and multiplied by a loading factor. These "weighted" scores are then totalled and divided by the number of T-units in the paper. A sample coding sheet showing the loading factors is given in Appendix O.

The scale developed by Golub, however, contains a

mathematical anomaly whereby an increase in the number of T-units analyzed decreases the syntactic density score independent of any other factor. Consequently, the mathematical procedures (but not the factors of syntactic density) were revised for the current study. A detailed discussion of the problems with the scale and the revisions used in this study are presented in Appendix D and also in an article written by the investigator in Research in the Teaching of English (Belanger, 1978).

i. procedures: prior to analyzing the papers, a coding sheet listing Golub's ten factors was duplicated and attached to each composition. The coding sheets provided space to "tic" each occurrence of a feature. The compositions were segmented into T-units and then items 2 to 4 (subordinate clauses per T-unit, mean main-clause word length, and mean subordinate-clause word length) were tabulated. The papers were then re-read and the grammatical features in items 5 to 10 were tabulated.

ii. training and check-coding: The investigator and a check-coder (Rater C, above) spent two two-hour sessions using Golub's (1973) "Self-Instructional Packet" (ED 095 538) to analyze the 15-composition sample of the papers used in the training sessions for the quality analysis noted above. Each coder had a copy of each paper and the analysis of each paper was discussed upon completion.

The check-coder then analyzed photocopies of 20

papers randomly chosen from the 288 compositions selected for syntactic density analysis. A discussion of the coding agreement between the coder and the check-coder is presented in Chapter IV.

d. T-unit length: O'Donnell's (1976) review of indices of syntactic maturity suggests that Hunt's (1965) T-unit has several virtues as a measure of growth in writing ability: it is simple, objective, and valid (p. 33). As defined by Hunt, T-units are the "shortest grammatically terminable units into which a connected discourse can be segmented without leaving any fragments as residue" (1965, p. 34). The T-unit, then, is one main clause with all the subordinate clauses attached to it or embedded within it. Hunt suggests that T-unit length is a more reliable indicator of linguistic maturity than sentence length because of the immature writer's tendency to link grammatically complete sentences together with co-ordinating conjunctions (and, but, so) while the more mature writer augments his sentences with more sophisticated linguistic expansions. Rules for segmenting the compositions into T-units are presented in Appendix L.

All 582 compositions collected in the study were examined for T-unit length. A coding sheet was not used here; instead, the papers were segmented into T-units on the papers themselves by marking the end of each T-unit in red ink. Since T-unit length is part of the syntactic density score, the check-coder's ratings were abstracted from the syntactic density check-coding and compared with the

investigator's ratings. This comparison is presented in Chapter IV.

e. fluency: R. Martin (1968), examining predictors of "good" tenth-grade writing, found that the length of composition contributed appreciably to discriminating between the good and poor writers: the good writers tended to write longer compositions than the poor writers did. Researchers such as Obenchain (1971), Palmer (1971), Miller and Ney (1968), Moore (1971), Zanotti (1970), and Harris (1962) also reported that the length of a student's composition was a good indicator of its quality. In the current study, the length of all 582 compositions in the writing sample was tabulated.

f. infrequency of major errors: The infrequency of major errors (fragments, run-on sentences, errors in subject-verb agreement, and errors in pronoun agreement) was a second predictor of "good" tenth-grade writing found by R. Martin (1968). Since these writing conventions were assumed to be taught (or induced over a long period of time) and since no specific writing instruction was offered as part of this experiment, significant changes in this area were not hypothesized. Nevertheless, a subsample of the papers (the 216 compositions used in the quality analysis) was analyzed for major errors. The results of this analysis are presented in Chapter V.

F. Statistical Procedures

Data collected in the study were analyzed using four

major statistical procedures: analysis of covariance, two-way analysis of variance, Pearson product-moment correlations, and t-tests. In all cases F-ratios and t-ratios at or beyond the .05 level of confidence were considered statistically significant.

1. Analysis of Covariance. In the current experiment the initial reading scores indicated that, although treatments were assigned randomly in each of the four pairs of classes, the experimental and control groups were not, in fact, of equivalent ability at the beginning of the experiment (see Chapter IV) and therefore analysis of covariance[1] was chosen as the statistical procedure. In this experiment the pretest for each measure was used as the covariate for the pre-mid and pre-post comparisons, and the midtest was used as the covariate for the mid-post comparison. That is, in the case of the pre-mid comparison, for example, the midtest scores of the experimental and control groups were adjusted statistically in light of differences between the two groups on the pretest so that the groups were statistically equivalent. In this instance, where the experimental and control groups were being compared only at the midtest, the midtest scores of the two groups had been adjusted to eliminate the statistical effect of initial differences between the two groups.

---

[1]The applications, assumptions, and procedures of this technique are discussed in Ferguson (1971).

The reading scores and the scores on each of the four major writing measures (quality, syntactic density, T-unit length, and fluency) of the experimental and control groups were compared by analysis of covariance at each of the three test comparisons (the pre-mid, the mid-post, and the pre-post comparisons). These comparisons were made at the University of Alberta Computing Center using the program ANOV 35: Three-Way Analysis of Variance and Covariance.

In addition, the program tested for interactions between the treatment and the sex of the subject and the treatment and the teacher who taught one of the four pairs of experimental and control classes in the experiment. These interactions indicate if the treatment (in this case an increase in reading ability) is more effective in altering the writing of the boys or the girls or if the methods used by one of the four teachers in the study were more or less effective for the treatment and non-treatment groups.

2. Two-Way Analysis of Variance. Two-way analysis of variance[1] was used to examine whether or not the treatment had differing effects for the high-, medium-, and low-ability students. The experimental and control groups were divided into the three ability levels on the basis of the pre-reading scores and then the groups were compared at each ability level on the reading and four writing measures at each of

---

[1]The applications, assumptions and procedures of this technique are discussed in Ferguson (1971).

the three test comparisons. These comparisons were made at the University of Alberta Computing Center using the program ANOV 35: Three-Way Analysis of Variance and Covariance.[1]

3. Pearson Product-Moment Correlations. Pearson Product-Moment Correlations are used to determine the relationship between the rankings of two sets of scores,[2] and results of the analysis are given both in terms of a correlation coefficient and an estimate of the statistical significance of the correlation. Since the statistical significance of the correlation relies heavily on the sizes of the groups being compared, and since the sizes of the groups in this study vary considerably depending upon the measure being discussed, correlational data are reported both in terms of the correlation coefficient and the statistical significance.

In the current study the major interest was in the relationship between reading ability and writing ability; consequently, correlations were computed between the reading scores and each of the four writing scores at each of the three test periods. In addition, correlations were computed separately for the boys and the girls and for the grade nines and grade tens. Finally, correlations among the four writing

---

[1] Analysis of variance and analysis of covariance may be used on the same computer program because the mathematical procedures are the same for both analyses: analysis of variance. The differences in the programs lie in the subroutines used prior to the analysis.

[2] The applications, assumptions, and procedures of this technique are discussed in Ferguson (1971).

measures were computed. Correlations were also used to determine the rater reliability reported in Chapter IV.

These computations were made at the Division of Educational Research Services, University of Alberta, using the program DEST05: Correlations with Optional t-tests.

4. t-Tests. The t-test[1] is a simple parametric statistic used to examine differences between two groups. Although less discriminating than the analysis of variance or analysis of covariance techniques, it has the advantage that it can be used readily for uncomplicated computations using a hand calculator rather than a computer. In the current experiment t-tests were used to determine the statistical probability of a number of comparisons which were tangential to the main analysis: the analysis of major errors, and the success of the randomization procedures, for example.

G. Summary

To examine the relationship between changes in reading ability and changes in writing ability, eight intact classes, two pairs of classes on the grade-nine level and two pairs on the grade-ten level, each pair taught by one teacher, were randomly designated experimental and control groups. Through drop-outs and absence during testing periods the original experimental population consisting of 112 experimental and 112 control students fell to 94 experimental

---

[1]The applications, assumptions, procedures, and conversion tables used for this technique are discussed in Ferguson (1971).

and 100 control students over the six months of the experiment.

Three forms of the Davis Reading Test, Series 2, were used as the reading measure and one in-class composition written at each test period was used as the writing sample. The composition topics "home," "school," and "society," the three most frequent choices of the students in Bell's (1971) survey, were each assigned to one-third of the sample at each of the three test periods.

Following administration of the initial reading and writing tests, the experimental groups were administered the S.O.S. Reading Technique, the treatment in the current study. Parallel reading and writing tests were administered three months and six months following the treatment. This design offered three possible test comparisons to examine the temporal relationship between the reading and writing scores: were the writing changes prompt (the first three months)? delayed (the second three months)? or spread evenly over the entire six months?

Four methods were used to analyze the compositions gathered in the study:

1. Quality Analysis. A 37-percent sample (216 of the 582 compositions) was analyzed for overall writing quality using a method suggested by Diederich (1974).
2. Syntactic Density Analysis. A 50-percent sample (288 compositions) was analyzed using a modified version of Golub's (1973) Syntactic Density Score.

3. _T-Unit Length_. All 582 compositions collected in the study were analyzed for average T-unit length following procedures suggested by Hunt (1965).
4. _Fluency_. The number of words in each composition was tabulated.

In addition, the major sentence errors in a sample of the compositions of the experimental and control students (the same sample used in the quality analysis above) were tabulated.

Because the initial measures indicated that the experimental and control groups were not of equivalent ability at the onset of the experiment, the statistical procedure of analysis of covariance was used to compare the experimental and control groups on the reading and four major writing measures in the study. Interactions between the treatment and the teacher who taught one of the four pairs of experimental and control classes and between the treatment and the sex of the subject were examined. In addition, two-way analysis of variance was used to examine the effects of the treatment on high-, mid-, and low-ability students, and correlations were computed between the reading and four writing measures used in the study.

CHAPTER IV

PRELIMINARY AND TANGENTIAL ANALYSES

Prior to proceeding with the main analysis of the data a number of preliminary analyses were carried out. First, it seemed advisable to examine in detail the reading changes shown by the experimental and control groups over the six months of the experiment. Although the reading experiment was a separate and complete study in itself (Martin and Belanger, 1977),[1] writing examination of the current study was predicated on differences in reading growth between the experimental and control groups.

Since intact classes were used in this experiment rather than completely randomized groups, the initial equivalence of the experimental and control groups on the reading and the four writing measures was examined. This analysis determined the statistical procedure to be used. Since the experimental and control groups were not equivalent on the reading and four writing measures at the beginning of the experiment, analysis of covariance was chosen as the appropriate statistic, as discussed in Chapter III.

---

[1]As noted in Chapter III, the sample sizes of the current study and the reading study vary slightly because 11 subjects who wrote all three reading tests did not write all three writing tests. The writing study had a sample size of 194 while the reading study had a sample size of 205.

Random samples consisting of 37 percent of the compositions collected in the study was analyzed for overall quality while a 50-percent random sample was used for the syntactic density analysis. As a check on the success of the sampling procedures the initial reading scores of the students in the samples were compared with the initial reading scores of the remainder of each group. This was done for the experimental and the control groups on both the grade-nine and grade-ten levels. As an additional check, these comparisons were also made on the pre-mid and pre-post reading gains scores of the sample groups and the remainder of the classes. Had the samples been totally unrepresentative of the groups as a whole this would have suggested the need for a subsidiary analysis of a different sample which, of course, could have been reported only as such.

Prior to the computer analysis of the data in the study an assessment was made of the varying difficulty of the composition topics assigned for the writing analysis. This was done for the grade-nine sample and the grade-ten sample separately. Since the three topics did not prove to be equally difficult, and since some topics were more difficult for grade-nine students than for grade-ten students, separate adjustment factors were calculated for the grade-nine and the grade-ten subsamples. These adjustment factors were calculated for each of the four writing measures and the score a given paper received on each measure was multiplied by the weighting factor for that particular topic before

computer calculations were made. In addition, eight of the thirty-six null hypotheses were tested for significance using the "raw" or unadjusted scores on the compositions and these computations are reported in this chapter.

Agreement among the three raters on the overall quality measure and agreement between the coder and the check-coder for the syntactic density and T-unit analyses were also examined.

Correlations among the four writing measures used in the study were computed, and comparisons of the T-unit data obtained in the current study with data from other studies were made.

A. Reading Analysis

When compared with the control students, the experimental students, those administered the reading treatment, showed statistically significant increases in reading ability on both the initial three-month comparison (the pre-mid comparison) and the six-month comparison (the pre-post comparison), but there was no significant difference between the two groups on the second three-month comparison (the mid-post comparison). That is, in effect, the experimental students gained significantly by the end of three months and simply maintained their superiority at the end of six months, without a further significant widening or narrowing of the gap demonstrated after three months.[1]

---

[1]To put the reading data in perspective, one can say

The pre-mid and pre-post differences were significant at or beyond the .002 level of confidence. When the grades were examined separately, the grade-ten experimental students showed reading gains superior to those of the control students at both the pre-mid and pre-post comparisons, but the grade-nine experimental students were superior to the control students on only the pre-post comparison. There were no significant interactions between the treatment and the sex of the subject or between the treatment and the teacher who taught one pair of experimental and control classes on any of the three test comparisons or on any of the three grade-level comparisons.[1, 2]

---

that the "gap" between experimentals and controls at the end of three months was approximately one-half year (2.85 raw score units) in favor of the experimentals. By a very narrow and nonsignificant margin, this gap had widened still further at the end of six months (2.91). The maintenance (indeed, slight widening) of this gap is contrary, of course, to what one might expect as the normal diminishing effect of instruction after the passage of considerable time, and this fact lent extra interest to the reading results. (Cf. also Martin and Belanger, 1977). Results were virtually identical for the reading portion of the analysis discussed here (n = 194) and the Martin and Belanger experiment (n = 205).

[1]As noted in Chapter III, the experimental and control groups were compared by three-way analysis of covariance using the treatment, teacher, and the sex of the subject as the criterion variables and the first test of each pair (for example, the pretest in the case of the pre-mid comparison) as the covariate.

[2]The statistical analysis of the reading scores and the reading gains scores for the three test comparisons and the three grade-level comparisons are reported here. The raw data by grade, treatment, sex, and test period are reported in Appendix G.

TABLE 3

Reading scores on the midtest using the pretest as the covariate. Probabilities are presented for the treatment effect, the teacher interaction, and the sex interaction for grades nine and ten separately and the two grades combined. The changes in the scores on the reading measure between the pre- and midtests are presented by treatment, grade, and sex.

	Grade 9		Grade 10		Grades Combined	
Statistical Data	n	prob	n	prob	n	prob
Treatment	103	0.16	91	0.004*	194	0.002*
Teacher Interaction	103	0.88	91	0.91	194	0.57
Sex Interaction	103	0.19	91	0.81	194	0.38
Reading Change						
Experimental	n	raw score	n	raw score	n	raw score
Boys	28	+0.39	25	+6.12	53	+3.09
Girls	22	-2.59	19	+4.00	41	+0.46
Combined	50	-0.92	44	+5.21	94	+1.95
Control						
Boys	24	-1.50	23	+1.13	47	-0.21
Girls	29	-2.66	24	-0.13	53	-1.51
Combined	53	-2.13	47	+0.49	100	-0.90

*$p < .05$

1. The Pre-Mid Comparison. As Table 3 indicates, when the pretest was used as the covariate, the grade-ten experimental groups and the grade-nine and -ten experimental groups combined were significantly different from their control counterparts on the midtest at the 0.004 and the 0.002 level of confidence respectively. On the other hand, the grade-nine

experimental and control groups were not significantly different on the pre-mid comparison.

Examination of the changes in the reading scores over this period (Table 3) indicates that the grade-ten experimental students gained more than five reading-score points while the control students gained less than one-half point over this period. When the two grades are combined, the experimental students show gains of almost two points while the control students show losses of almost one point. Examining the grade-nine students separately shows that both the experimental and control groups experienced losses in reading scores over the initial three months but that the controls showed greater losses than the experimentals did. The grade-nine experimental students showed losses of almost one point over the three-month period while the grade-nine control students lost over two points during this time.[1] These differences, however, were

---

[1]The reading scores were calculated using Davis's (1962) "adjusted scores" (adjusted by subtracting one-quarter point for each incorrect answer from the total number of correct answers), a measure to correct for guessing. However, these are not Davis's normalized scores. Normalizing the scores would adjust for the differences in the difficulty of the three forms of the test, and thus eliminate the anomaly of students "losing" reading ability over the testing periods; however, the normalized scores are given in percentiles and, consequently, may not be averaged.

Therefore, the apparent losses experienced by some groups might be attributed to the varying difficulty of the three forms of the test. On the other hand, this does not account for all of the losses: all of the grade-ten groups showed appreciable losses between the mid- and posttests (see Table 4, column 2) which, at least in part, is the consequence of the apparent attitude of the senior high school students at the beginning of June.

not statistically significant.

TABLE 4

Reading scores on the posttest using the midtest as the covariate. Probabilities are presented for the treatment effect, the teacher interaction, and the sex interaction for grades nine and ten separately and the two grades combined. The changes in the scores on the reading measure between the mid- and posttests are presented by treatment, grade, and sex.

	Grade 9		Grade 10		Grades Combined	
Statistical Data	n	prob	n	prob	n	prob
Treatment	103	0.11	91	0.27	194	0.68
Teacher Interaction	103	0.08	91	0.96	194	0.09
Sex Interaction	103	0.51	91	0.73	194	0.72
Reading Change						
Experimental	n	raw score	n	raw score	n	raw score
Boys	28	+0.43	25	-2.52	53	-0.96
Girls	22	+6.05	19	-3.00	41	+1.85
Combined	50	+2.90	44	-2.73	94	+0.27
Control						
Boys	24	+1.63	23	-2.74	47	-0.51
Girls	29	+1.62	24	-0.13	53	+0.83
Combined	53	+1.62	47	-1.40	100	+0.20

2. The Mid-Post Comparison. As Table 4 indicates, when the midtest was used as the covariate, there were no significant differences between the experimental and control students over the second three-month period of the experiment. During this

time, the reading gains shown by the grade-nine and -ten classes combined were almost equal for the experimental and control groups. On the other hand, when the grade-nine and -ten classes were examined separately, the grade-nine students, who showed losses over the pre-mid comparison, showed reading gains, the experimentals gaining almost twice as much as the controls, while the grade-ten students, who showed gains over the pre-mid comparison, showed reading losses, again the experimentals almost twice as much as the controls.

3. The Pre-Post Comparison. As Table 5 shows, when the pretest was used as the covariate, the experimental and control groups were significantly different on the posttest, both when the grade-nine and -ten students were examined separately and when the two grades were combined. Over the six-month period, the grade-nine experimental students gained almost two reading-score points while their control counterparts lost almost one-half point. At the same time, the grade-ten experimental students gained almost two and one-half points while their control counterparts lost almost a full point.

Comparing the pre-post gains scores with Davis's (1962) normalized data indicates that the experimental students on the grade-nine level, who gained 2.49 points more than their control counterparts, showed 0.41 of a year's growth more than the control students.[1] The grade-ten

---

[1]A "year" on the Davis norms is a rough approximation and varies with the grades involved and with the levels of the

TABLE 5

Reading scores on the posttest using the pretest as the covariate. Probabilities are presented for the treatment effect, the teacher interaction, and the sex interaction for grades nine and ten separately and the two grades combined. The changes in the scores on the reading measure between the pre- and posttests are presented by treatment, grade, and sex.

	Grade 9		Grade 10		Grades Combined	
Statistical Data	n	prob	n	prob	n	prob
Treatment	103	0.020*	91	0.023*	194	0.0013*
Teacher Interaction	103	0.16	91	0.69	194	0.55
Sex Interaction	103	0.93	91	0.48	194	0.65
Reading Change						
Experimental	n	raw score	n	raw score	n	raw score
Boys	28	+0.82	25	+3.60	53	+2.13
Girls	22	+3.46	19	+1.00	41	+2.32
Combined	50	+1.98	44	+2.48	94	+2.21
Control						
Boys	24	+0.13	23	-1.61	47	-0.72
Girls	29	-1.03	24	-0.25	53	-0.68
Combined	53	-0.51	47	-0.92	100	-0.70

* $p < .05$

experimental students showed somewhat greater gains over the control students, a difference of 3.4 points, or 0.57 of a

---

raw scores. As a general rule, about six raw-score units per year appears to be the difference between the grade levels on Davis's norms.

year's reading growth.

B. Initial Equivalence of Groups

The use of intact classes rather than fully randomized groups required an assessment of the initial equivalence of the experimental and control groups on the reading and the four writing measures. Comparison of the experimental and control groups by sex and by grade was done by means of t-tests.

Although there were initial differences between the experimental and control groups on all five measures examined, the differences were statistically significant on only two of the measures, the reading and the T-unit length measures. On the reading measure, the grade-nine experimental boys were substantially below their control counterparts, while on the T-unit length measure the grade-ten experimental girls were substantially above the control girls. Both of these differences were significant beyond the .05 level of confidence. These differences suggested that analysis of covariance was the appropriate statistical procedure for the major analyses. The data for these analyses are presented and discussed in Appendix M.

C. Success of Sampling Procedures

As discussed in Chapter III, a 37-percent sample of the papers gathered in the study was analyzed for overall quality while a 50-percent sample was analyzed for syntactic density. Since it was possible that these samples were not

representative of the experimental and control groups as a whole, the students included in the sample groups were compared with the students not included in the sample groups to give an estimate of the representativeness of the samples. t-tests were used to examine the statistical significance of the differences between the sample groups and the remainder of the experimental population on the initial reading test, on the three-month reading-gains scores, and on the six-month reading-gains scores. Comparisons were made for the experimental and control groups, for the boys and the girls separately, for the grade-nine and grade-ten students separately, and for the grade-nine and grade-ten groups combined.

Although there were differences in scores between the samples and the remainder of the experimental populations on all three variables, these differences reached statistical significance at the .05 level of confidence on only one comparison: the reading-gains scores over the six months of the experiment for the random sample of experimental students selected for the syntactic density analysis were significantly different ($p = .01$) from those of the remainder of the experimental population. Since the differences between the samples and the remainder of the experimental populations were, for the most part, both statistically non-significant and in the same direction for the experimental and control samples, these differences were not expected to influence the results of either the quality analysis or the syntactic

density analysis. The data for these analyses are presented and discussed in Appendix N.

D. Adjustment for Topic Difficulty

Provision was made in the original project proposal to adjust the scores on the three composition topics because it was felt that the topics might not prove to be equally difficult for the students. This inequality of topic difficulty was confirmed both by raters on the quality analysis who suggested that the students appeared to perform better on the "school" and "home" topics than they did on the less-immediate topic "society," and by the tabulations of the scores by topic on each measure. Since the grade-nine students and the grade-ten students in the study received different average scores on the topics (the grade-ten students, for example, scored much higher on the topic "society" than the grade-nine students did), separate computations were made for each grade.

The average score for each topic on each measure and at each grade level having been found, the topic receiving the highest average score was chosen and the average of the scores for the other two topics were subtracted from it. These remainders were expressed as a percentage of the topics receiving the highest score and the scores on the other topics were multiplied by one plus this percentage to equate them with the topic receiving the highest score. These calculations are presented in Appendix E.

1. Grade-Nine Adjustments. On three measures, syntactic

density, T-unit length, and fluency, grade-nine students scored the highest on Topic J (School). Therefore, scores on the other two topics (K, Home; and L, Society) were multiplied by the loading factor in Table 6 to equate them with Topic J. On one measure, quality analysis, Topic K, (Home) received the highest scores on the grade-nine level and, consequently, Topic J (School) and Topic L (Society) were multiplied by the loading factor in Table 6 to equate them with Topic K.

2. Grade-Ten Adjustments. On the grade-ten level, scores for Topic J (School) on the quality, T-unit length, and fluency measures were higher than those for Topics K (Home) and L (Society). For the syntactic density analysis, on the other hand, Topic L (Society) received the highest scores. These topics were equated by multiplying them by the loading factors given in Table 7.

E. Comparison of Selected Adjusted and Unadjusted Scores

As noted above, the three topics for the compositions written in the study, "home," "school," and "society," were not of equal difficulty, as evidenced by the average scores obtained on each of the three topics. Consequently, prior to the statistical analysis each composition was multiplied by a loading factor for each of the four composition measures.

To examine the influence of these adjustments, eight of the thirty-six null hypotheses tested in the study were recomputed using the unadjusted scores. Comparison of the

TABLE 6

Means, standard deviations, and loading factors for grade-nine compositions on each of the four writing measures for each of the three composition topics

	Topic J				Topic K				Topic L			
Grade 9	n	$\bar{x}$	σ	Loading Factor	n	$\bar{x}$	σ	Loading Factor	n	$\bar{x}$	σ	Loading Factor
Quality	36	5.21	1.34	1.085325	36	5.65	1.44	1.0	36	5.08	1.35	1.112014
Syntactic Density	48	3.59	0.77	1.0	48	3.50	0.93	1.07218	48	3.44	0.73	1.04391
T-Units	103	14.37	2.44	1.0	103	14.36	3.30	1.00091	103	14.10	2.83	1.01893
Fluency	103	245.95	85.71	1.0	103	239.04	92.49	1.02892	103	236.17	85.07	1.04139

TABLE 7

Means, standard deviations, and loading factors for grade-ten compositions on each of the four writing measures for each of the three composition topics

Grade 10	Topic J n	Topic J $\bar{x}$	Topic J σ	Topic J Loading Factor	Topic K n	Topic K $\bar{x}$	Topic K σ	Topic K Loading Factor	Topic L n	Topic L $\bar{x}$	Topic L σ	Topic L Loading Factor
Quality	36	5.84	1.78	1.0	36	5.79	1.37	1.009496	36	5.66	1.79	1.031760
Syntactic Density	48	3.84	1.08	1.005200	48	3.59	0.83	1.076660	48	3.86	0.86	1.0
T-Units	91	15.23	3.52	1.0	92*	14.76	2.89	1.03246	90	14.94	3.26	1.01954
Fluency	91	272.29	99.96	1.0	92*	249.80	98.89	1.090043	90	248.96	88.30	1.09376

*The numbers of compositions written on Topics K and L are not equal because one student, a grade-ten experimental boy, writing a make-up test in June insisted that he had written on Topic L previously, but not on Topic K. Consequently he wrote on Topic K for both the pre- and posttest. Since he had written on Topic K as a pretest six months previously, it was thought that this doubling would not materially affect the results, and the student's papers were included in the analysis.

analyses obtained from the adjusted and unadjusted scores for the main effects, the teacher-treatment interaction and the sex-treatment interaction are presented in Tables 8, 9, and 10 respectively.

TABLE 8

The effect of adjustment for topic difficulty on the treatment variable for grades nine and ten combined. The statistical data for analyses of scores adjusted for topic difficulty and scores not adjusted for topic difficulty are presented for a sample of 8 of the 36 null hypotheses tested in the study.

Measure	Time	Adjustment	D.F.	F Ratio	Prob	Decision
Quality	Pre-Mid	No	1	0.636	.426	N.S.
		Yes	1	0.001	.970	N.S.
Quality	Mid-Post	No	1	0.062	.804	N.S.
		Yes	1	0.0002	.963	N.S.
Quality	Pre-Post	No	1	0.218	.641	N.S.
		Yes	1	0.030	.862	N.S.
Fluency	Pre-Mid	No	1	1.687	.196	N.S.
		Yes	1	1.714	.192	N.S.
Fluency	Mid-Post	No	1	4.206	.042	Sig*
		Yes	1	4.337	.039	Sig*
Fluency	Pre-Post	No	1	0.001	.974	N.S.
		Yes	1	0.0003	.986	N.S.
T-Units	Pre-Post	No	1	0.429	.513	N.S.
		Yes	1	0.478	.490	N.S.
Syntactic Density	Pre-Post	No	1	0.192	.662	N.S.
		Yes	1	1.309	.256	N.S.

*$p < .05$

As Table 8 shows, when the main effects were examined, the adjustment did alter the statistical probability of the differences between the experimental and control groups.

However, the adjustment did not alter any decision to accept or reject the null hypothesis on any measure. At the .05 level of confidence, the experimental and control groups were significantly different on only one of the eight measures examined, the fluency measure on the pre-post comparison, and this decision was the same using either the adjusted or unadjusted scores. A similar pattern is observed for both the

TABLE 9

The effect of adjustment for topic difficulty on the interaction between the teacher and the treatment variables for grades nine and ten combined. The statistical data for scores adjusted for topic difficulty and scores unadjusted for topic difficulty are presented for a sample of 8 of the 36 null hypotheses tested in the study.

Measure	Time	Adjustment	D.F.	F Ratio	Prob	Decision
Quality	Pre-Mid	No	3	1.608	.189	N.S.
		Yes	3	0.832	.482	N.S.
Quality	Mid-Post	No	3	1.166	.324	N.S.
		Yes	3	0.521	.669	N.S.
Quality	Pre-Post	No	3	0.407	.748	N.S.
		Yes	3	0.856	.469	N.S.
Fluency	Pre-Mid	No	3	0.592	.621	N.S.
		Yes	3	0.469	.705	N.S.
Fluency	Mid-Post	No	3	0.702	.552	N.S.
		Yes	3	0.763	.516	N.S.
Fluency	Pre-Post	No	3	0.559	.643	N.S.
		Yes	3	0.493	.687	N.S.
T-Units	Pre-Post	No	3	1.178	.319	N.S.
		Yes	3	1.198	.312	N.S.
Syntactic Density	Pre-Post	No	3	1.055	.370	N.S.
		Yes	3	0.680	.567	N.S.

treatment-teacher interaction presented in Table 9, and the sex-treatment interaction presented in Table 10. As Table 9 shows, at the .05 level of confidence there were no significant interactions between the treatment and the teacher who taught one of the four pairs of experimental and control classes either when the adjusted or unadjusted scores were examined. As Table 10 shows, only one of the interactions between the

TABLE 10

The effect of adjustment for topic difficulty on the interaction between the sex of the subject and the treatment variables for grades nine and ten combined. The statistical data for scores adjusted for topic difficulty and scores not adjusted for topic difficulty are presented for a sample of 8 of the 36 null hypotheses tested in the study.

Measure	Time	Adjustment	D.F.	F Ratio	Prob	Decision
Quality	Pre-Mid	No	1	0.036	.850	N.S.
		Yes	1	0.228	.635	N.S.
Quality	Mid-Post	No	1	0.513	.475	N.S.
		Yes	1	0.012	.912	N.S.
Quality	Pre-Post	No	1	0.477	.491	N.S.
		Yes	1	0.006	.940	N.S.
Fluency	Pre-Mid	No	1	1.137	.288	N.S.
		Yes	1	0.848	.358	N.S.
Fluency	Mid-Post	No	1	1.290	.258	N.S.
		Yes	1	1.852	.175	N.S.
Fluency	Pre-Post	No	1	2.268	.134	N.S.
		Yes	1	2.255	.135	N.S.
T-Units	Pre-Post	No	1	4.433	.037	Sig.*
		Yes	1	4.384	.038	Sig.*
Syntactic Density	Pre-Post	No	1	2.628	.107	N.S.
		Yes	1	1.394	.241	N.S.

*$p < .05$

treatment and the sex of the subject, the interaction for the pre-post T-unit comparison, was significant at the .05 level of confidence, and this decision was the same using either the adjusted or unadjusted scores.

F. Rater Agreement

Computations were made on the agreement among the three raters for the quality analysis and between the coder and the check-coder for the syntactic density analysis and the T-unit analysis.

1. Rater Agreement: Quality Analysis. As Table 11 shows, the two primary raters, raters A and B, assigned exactly the same score to 68 of the 216 papers or 31.48 percent of the papers rated in the quality analysis. They were not more than one grade apart on an additional 80 papers (37.04 percent). This left the arbitrator, rater C, 68 papers.[1] Of these 68 papers, he assigned the same score as rater A or B to 26 papers (38.24 percent) and was within one grade of rater A or B on an additional 35 papers (51.47 percent). Two of the three raters, then, were not more than one grade apart on 209 of the 216 papers rated in the study (96.76 percent). The scores of the remaining seven papers were calculated using the method suggested by Diederich (1974): the grade given by rater C was averaged with the grade of the other two raters

---

[1]Diederich suggests that about ten percent of the papers will require rerating by rater C. However, Diederich is using the five-point letter grade system, a somewhat grosser measure than the eleven-point scale used here. Therefore, it could be expected that a larger number of papers than he suggests would require rerating.

TABLE 11

Agreement among Raters A, B, and C on the quality measure presented in terms of both the total agreement and percentage of agreement.

	Number of Compositions	Percentage of Compositions
Raters A and B assigned the same score	68/216	31.48
Scores of Raters A and B within one point of each other	80/216	37.04
Overall agreement Raters A and B	148/216	68.52
Compositions scored by Rater C	68/216	31.48
Compositions assigned the same score by Rater C as either Rater A or Rater B	26/68	38.24
Score of Rater C within one point of score of Rater A or Rater B	35/68	51.47
Score of Rater C more than one point away from scores of both Rater A and Rater B	7/68	10.29
Total Agreement among Raters A, B, and C	209/216	96.76

which was farthest from the mean of the sample. A discussion of the rater agreement by teacher and grade as well as an explanation of the seven papers that the raters did not agree on is presented in Appendix H.

The method of rating papers suggested by Diederich (1974) and used in the current study does not lend itself to the traditional methods of calculating interrater reliability.

According to Diederich (1974) the correlations are computed on the "two ratings that actually determine the grade" (p. 35) and consequently appropriate ratings of rater C were substituted before Pearson product-moment correlations were computed for rater reliability. Traditionally, of course, the ratings of all raters are pooled for the reliability analysis.

TABLE 12

Rater reliabilities by day and the number of raters. (Pearson product-moment correlations).

Day	Three Raters	Two Raters
1	.868	.458
2	.881	.689
Grade 10	.868	.581
3	.847	.398
4	.876	.430
Grade 9	.859	.433
Overall	.858	.506

As Table 12 shows, the overall rater reliability for the quality analysis was .858 which is considerably greater than the .67 reliability Diederich (1974) suggests can be expected from using this method to mark high-school essay examinations. In fact, it is even higher than the .80 he suggests can be obtained from collecting two essays from each student and pooling the scores on the essays (p. 35). On the other hand, the reliability figure reported here

obtained using the holistic method of rating is somewhat below the .88 and .91 reliabilities for pre- and posttests which Buxton (1959) obtained with two raters using a very detailed rating scale. Rating scales, however, have not produced consistent results. Fosveldt (1962), for example, attempting to validate a number of composition rating scales, did not compute reliabilities because he found that the only thing that his raters agreed upon was the merit of the items in the scales.

The holistic method of rating has been shown to produce high rater reliabilities. Wiseman (1949), for example, found reliabilities of .90 between four judges who each rated 50 to 70 papers per hour,[1] but the task of Wiseman's raters was merely to "assess whether a student could profit from a grammar school education," and consequently the papers were graded on a pass-fail basis. Researchers providing little instruction for the raters have reported low reliabilities. Diederich, French, and Carlton (1961), for example, found reliabilities of .31 among 54 raters with diverse backgrounds and of .41 among 9 English teachers.

Other researchers have reported high reliabilities, but the methods of calculating the quotients were unconventional. Giffen (1972) reported rater reliabilities of .73 among three raters. However, Giffen used a different method of calculating the reliability quotients than any of

---

[1]The judges in the current study rated approximately 27 papers each per hour.

the above studies did[1] and consequently the results may not be comparable. Gajadharsingh (1970), using a four-point scale and again using no formal program of rater training, reported reliabilities of .846 to 1.00 among the three raters but he used an unconventional method of obtaining the reliabilities.[2]

Table 12 also offers a comparison of the reliabilities obtained using Diederich's three-rater method with the reliabilities obtained using the two primary raters, raters A and B, alone. It can be seen that Diederich's method not only brings the rater reliabilities to an acceptable level but also tends to minimize the daily fluctuations. Table 12 reports the rater reliabilities chronologically by the day that the papers were scored: on the first two days grade-ten papers were scored and on the last two days grade-nine papers were scored. In informal discussions following the sessions, the raters suggested that they felt that they were more competent on the second day of rating than the first. This, despite the fairly protracted rater training sessions, was

---

[1]Giffen chose to treat the two writing assignments collected in his study as parallel tests, calculated differences among the raters through analysis of covariance, and translated this to a reliability quotient using a generalizability formula.

[2]Instead of calculating the reliability of compositions in the study, the traditional method, Gajadharsingh gave his raters five compositions to rate prior to their rating of the sample. He calculated the reliability quotients on these five papers alone. Since three of these papers were written by professional writers and two by grade-nine students, and since the four-point scale was very general, the high correlations he obtained are not surprising.

evident in the two-rater reliabilities on the first two days, .458 and .689 respectively. They also suggested that they found the grade-nine papers more difficult to rate than the grade-ten papers and this again is evident in the two-rater reliability figures. However, when the rating of the third rater is substituted for that of the other rater which is farthest away from the third rater's mark, these daily fluctuations are minimized.

2. Agreement Between Coder and Check-Coder. Following the training sessions discussed in Chapter III, the check-coder independently analyzed photocopies of 20 of the compositions to be used in the writing analysis for syntactic density. Since T-unit length is a part of the Syntactic Density Score, this tabulation was abstracted from the Syntactic Density Score for comparison with the investigator's findings.

a. Syntactic Density Score. Although the coder and the check-coder arrived at exactly the same score on only one of the 20 papers, the differences between the codings on the other papers were not substantial. The average difference between the coder and check-coder's ratings, as shown in Table 13, was 0.014 points (σ 0.072; range 0.25 to -0.07). Since the average Syntactic Density Score for the compositions examined in the study was 3.731 (σ 0.905) this relatively minor difference of 0.004 percent in the two codings should not affect the results of the experiment.

b. T-unit length. As Table 14 shows, the coder and the

TABLE 13

Agreement between coder and check-coder on the Syntactic Density Score.

Composition Number	Coder	Check-Coder	Coder Minus Check-Coder
60	3.94	4.01	-0.07
135	2.21	2.20	+0.01
113	3.94	4.00	-0.06
145	3.66	3.72	-0.06
210	2.78	2.84	-0.06
182	3.55	3.53	+0.02
177	4.11	4.10	+0.01
246	3.03	3.05	-0.02
279	4.55	4.54	+0.01
254	3.51	3.58	-0.07
231	5.23	5.24	-0.01
*167	3.87	3.62	+0.25
74	4.51	4.54	-0.03
96	3.73	3.74	-0.01
103	5.51	5.51	0
23	4.47	4.57	-0.10
35	2.74	2.70	+0.04
17	3.90	3.95	-0.05
7	3.69	3.75	-0.06
120	2.82	2.84	-0.02

*Coder and check-coder disagreed on the number of T-units in the composition.
Range of differences (coder's score minus check-coder's score): +0.25 to -0.10 (mean -0.014; standard deviation 0.072).

check-coder disagreed on the number of T-units found in a paper on only one of the 20 papers compared. This composition, number 167, contained the sentence "I have always thought that school was made on a five day schedule so we could 'recouperate' (roughly speaking) from the week's work." In the training sessions the coders agreed that independent clauses joined by "so" would be counted as two T-units unless "so that" was very strongly implied. The coder felt that "so that" was very strongly implied and counted the above sentence as one T-unit. The check-coder counted it as two.

TABLE 14

Agreement Between the Coder and Check-Coder on the T-Unit Analysis

Composition Number	Coder: Words	Coder: T-Units	Coder: Words per T-Unit	Check-Coder: Words	Check-Coder: T-Units	Check-Coder: Words per T-Unit	Words: Coder Minus Check-Coder
60	136	9	15.11	135	9	15.00	+1
135	167	17	9.82	165	17	9.71	+2
113	326	22	14.82	325	22	14.77	+1
145	257	17	15.12	256	17	15.06	+1
210	88	7	12.57	85	7	12.14	+3
182	197	15	13.13	195	15	13.00	+2
177	162	10	16.20	159	10	15.90	+3
246	300	21	14.29	298	21	14.19	+2
279	159	9	17.67	157	9	17.44	+2
254	234	17	13.76	231	17	13.59	+3
231	265	13	20.38	264	13	20.31	+1
*167	277	18	15.39	276	19	14.53	+1
74	191	11	17.36	195	11	17.73	-4
96	158	11	14.36	159	11	14.45	-1
103	262	14	18.71	261	14	18.64	+1
23	227	21	10.81	230	21	10.95	-3
35	182	17	10.71	189	17	11.12	-7
17	265	17	15.59	265	17	15.59	0
7	241	17	14.18	246	17	14.70	-5
120	84	8	10.50	85	8	10.63	-1
Total	4178	291		4176	292		

*Coder and check-coder disagreed on the number of T-units.

However, as Table 14 shows, the coders did not agree on the number of words in the papers. The paper with the widest disagreement, paper number 241, had five words written on the back of the page which the coder failed to count but the check-coder counted. Other differences arose because one coder counted "a.m." as two words but the other counted it as one and because one coder counted the percent sign and the dollar sign as a word and the other did not. These differences, however, amounted to three-percent at the most with scarcely a perceptible influence on either the T-unit analysis or the fluency analysis.

G. Correlations Among Writing Measures

As Table 15 [1] shows, when the writing measures for the entire sample are correlated, the quality and fluency measures correlate very highly with each other and the syntactic density and T-unit length measures correlate very highly with each other; however, the syntactic density and quality, the syntactic density and fluency, the T-unit length and quality, and the T-unit length and fluency measures correlate only slightly with each other and in many cases the correlation is negative. The highest correlations

---

[1]Since the sample sizes used for the four measures vary widely and since the statistical significance of a correlation depends heavily on the sample size, both the correlation coefficient and the statistical significance of the correlation are reported in Table 15.

TABLE 15

Correlations among the four writing measures used in the study. Since the sample sizes vary from 72 to 194, both the correlation coefficients (r) and the statistical probabilities are given.

		Quality			Syntactic Density			T-Unit Length		
		n	r	prob	n	r	prob	n	r	prob
Syntactic Density	Pre	72	-0.069	0.562						
	Mid	72	-0.029	0.809						
	Post	72	-0.020	0.865						
T-Unit Length	Pre	72	-0.072	0.547	96	0.944	0.001			
	Mid	72	-0.059	0.623	96	0.901	0.001			
	Post	72	-0.031	0.797	96	0.904	0.001			
Fluency	Pre	72	0.347	0.003	96	-0.068	0.508	194	0.009	0.903
	Mid	72	0.483	0.0001	96	-0.255	0.012	194	-0.174	0.015
	Post	72	0.425	0.0001	96	0.087	0.399	194	0.094	0.190

observed, correlations of beyond 0.9 between T-unit length and syntactic density, support findings of O'Donnell (1976) who suggests that the measures are so closely related (because syntactic density relies heavily on T-unit length) that the extra work required by the syntactic density analysis does not appear justified. Even with the modified scale used here but not by O'Donnell the two measures appear to yield virtually the same kinds of results.

When the boys and girls in the study are examined separately and when the experimental and control groups are examined separately, the correlations are similar to those reported above. These tables are presented in Appendix I.

When the four writing measures were selected for the study they were selected on the basis of their measuring different aspects of composition quality. The quality and fluency measures, although closely related, appear not to be measuring exactly the same skills, and, therefore, using the two measures instead of one appears to be justified. The syntactic density and T-unit measures, on the other hand, appear to be measuring almost exactly the same skills, and, therefore, the value of using two measures instead of one appears questionable.

H. Comparison of T-Unit Data in the Current Study with Data from Previous Studies

As Table 16 shows, previous researchers examining writing on the grade-ten level reported average T-unit lengths ranging from 10.44 to 16.0 while grade-ten T-unit lengths

TABLE 16

T-unit lengths found in the current study and T-unit lengths found in previous studies.

	Grade 8	Grade 9	Grade 10	Grade 12
Current Study				
pretest		14.25	15.23	
midtest		14.22	15.53	
posttest		14.65	14.95	
Previous Studies				
Hunt (1965)	11.5			14.4
Hunt (1970)			10.44	
Potter (1966)				
good writers			16.0	
poor writers			14.2	
R. Martin (1968)				
good writers			12.67	
poor writers			12.69	

reported in the current study range from 14.95 to 15.53. The average T-unit length reported, of course, depends both on the topic the papers are written on and the rules for segmentation that the investigator uses.

The influence of topic is best illustrated by comparison of Hunt's 1965 and 1970 reports. In 1965 Hunt collected essays on topics randomly assigned by the teacher and reported that in this sample the average T-unit length on the grade-eight level was 11.5 while on the grade-twelve level it was 14.4, an average yearly increment of close to one word per T-unit. In Hunt's 1970 study however, where students were required to rewrite a passage on the topic Aluminum, the average T-unit length on the grade-ten level

was 10.44, over one word per T-unit shorter than those reported on the grade-eight level in his earlier study. Of course, not only the topic but the skill required would influence the findings here: the former was free writing while the latter was rewriting: combining a number of very short clauses into longer and more complex sentences. Nonetheless, Potter (1966), who used a topic very similar to those in the current study (his topic was "The Qualities of a Good Teacher"), reported findings closest to those in the current study. On the other hand, R. Martin's (1968) students wrote on the topic "How to Get from My Home to Dreher High School," an instructional topic which might be expected to elicit somewhat shorter T-units than the expository topics used in the current study would.

A second important factor in the average T-unit length reported in various studies, as noted in Chapter III, is the rules the investigator uses to segment the discourse into T-units. If one chooses, as Mellon (1969) did, to count clauses of condition, concession, reason, and purpose as independent T-units, the average T-unit length would clearly be somewhat shorter. The current study did not use Mellon's method, but instead used the method outlined in Chapter III.

However, in one respect the differences between the grade-nine and grade-ten students in the current study[1] are

---

[1]The grade-nine and grade-ten students, of course,

similar to the normative data reported by Hunt (1965): there is an average of just under one word per T-unit between the two grades.

I. Summary

This chapter reported the results of a number of preliminary analyses and analyses that were tangential to the investigation.

On the reading measure, statistically significant differences between the experimental and control groups were found on both the pre-mid and pre-post comparisons. On the pre-mid and pre-post comparisons both the grade-ten and the combined grade-nine and grade-ten experimental groups showed significant increases in reading ability ($p < .02$ and $.002$, respectively). The grade-nine experimental group showed significant gains ($p < .02$) on the pre-post comparison only. On the mid-post comparison, however, there were no significant differences between the experimental and control groups on the reading measure.

An examination of the initial equivalence of the experimental and control groups on the reading and the four writing measures showed statistically significant ($p < .05$) differences between the two groups on two of the five measures, the reading and the T-unit measures. Consequently, analysis

---

are sampled from two different populations. Because of the areas of the city in which the three schools are located and because of attrition rates, the grade-ten sample is not simply the grade-nine sample one year older. They are different populations.

of covariance was chosen as the statistical procedure for the major analyses.

The samples chosen for the quality and syntactic density analyses were found to be representative of the experimental population as a whole. Only one of the thirty-six comparisons made revealed differences significant at the .05 level of confidence.

Since the three composition topics, "home," "school," and "society," did not prove to be equally difficult for the students in the experiment, the score that each student received on each measure was adjusted to reflect the difficulty of the topic. A sample of eight of the thirty-six null hypotheses tested in the study was re-examined using the unadjusted scores, and the results obtained using the adjusted and unadjusted scores were compared. The comparison suggested that adjusting the scores for topic difficulty did not significantly influence the findings of the study.

Interrater reliability on the quality measure was .858 for the entire sample. It was noted that Diederich's (1974) suggested method of rating the compositions not only brought the interrater reliability to an acceptable level but also minimized daily fluctuations in ratings.

A sample of the compositions used in the syntactic density and T-unit analyses was analyzed by a check-coder. Disagreements between the coder and check-coder on these two measures were minor and were not expected to influence the results of the experiment.

Correlation coefficients above .90 were found between the T-unit and syntactic density measures and coefficients ranging between .35 and .48 were found between the quality and fluency measures. However, only slight correlations were found between the T-unit measure and the quality and fluency measures and between the syntactic density measure and the quality and fluency measures.

CHAPTER V

FINDINGS OF THE STUDY

The compositions of the experimental and control groups in the study were compared on four measures of writing skill: quality, syntactic density, T-unit length, and fluency. A sample of the compositions was examined for major sentence errors. In addition, correlations between the reading measure and the four composition measures were calculated and the average scores for each group on each measure were compared.

When the compositions of the experimental and control groups in the study were compared using analysis of covariance on the four major writing measures, statistically significant differences between the two groups were found on only three of the thirty-six comparisons. Only random significant interactions between the treatment and the sex of the subject or the treatment and the teacher who taught one of the four pairs of experimental and control classes were observed. When the experimental and control groups were compared on the four major writing measures by two-way analysis of variance to examine the effects of the treatment on high-, medium-, and low-ability students, no statistically significant differences were found.

A subsidiary analysis of major errors found in the compositions showed significant differences between the

experimental and control groups on only one of the nine comparisons made. However, as noted in Chapter III, significant differences were not hypothesized on this measure.

Comparison of the scores on the reading measure with the scores on the four writing measures suggests two different conclusions, depending on the method of comparison. Correlational data show a statistically significant relationship between the reading measure and two of the writing measures, the quality measure and the fluency measure. On the other hand, when the mean scores on the reading and four writing measures are compared by grade, treatment, and sex, no consistent pattern of relationships between the measures emerges.[1]

---

[1]As noted above, only three of the thirty-six null hypotheses tested in the study were rejected at the .05 level of confidence. All three of these differences favored the control group. This is approximately the number that would be expected to be rejected by chance alone, and, consequently, these results must be regarded as spurious.

The data in the present chapter offers absolutely no support for the hypothesis that changes in reading skill will be accompanied by changes in writing skill. While the correlations between the reading measure and two of the writing measures used in the study suggest some relationship between reading and writing skill, the major analyses fail to show any relationship between the two skills. On the major analyses, even the statistically non-significant differences are not consistently in the hypothesized direction. Some measures favor the experimental group while others favor the control group, with no apparent pattern of changes emerging.

Because of the paucity of evidence supporting the hypotheses of the investigation, the findings in this chapter are not presented as an examination of the null hypotheses. To the investigator, an unabridged account of the examination of the thirty-six null hypotheses, one by one, when the main result was clearly not in support of the hypothesized effect, appeared likely to be merely tedious and unenlightening as a part of the main exposition. Consequently, only summaries of

A. Analysis of the Four Writing Measures

The experimental and control groups were compared on the quality, syntactic density, T-unit length, and fluency measures by analysis of covariance, and the interactions between the treatment and the sex of the subject and between the treatment and the teacher who taught one of the four pairs of experimental and control classes were examined. As an aid to interpreting the statistical data, the writing change scores are presented for each writing measure by treatment, grade, and sex. The changes on the writing measures were hypothesized to follow the changes on the reading measure reported in Chapter IV. For convenience, the statistical probabilities of the changes on the reading measure reported in Chapter IV are included with the tables on the writing measures.[1]

1. Quality Analysis. Nine of the thirty-six null hypotheses tested in the study examined differences in the quality of the

---

the findings on the four major writing measures are presented in this chapter. For the reader who wishes to examine the data in detail, a discussion of the null hypotheses and the accompanying statistical data is presented in Appendix B.

[1]As was noted in Chapter III, the quality and syntactic density analyses are based on subsamples of the experimental and control groups. The reading analysis, however, is based on the entire experimental population. While the examination of sampling procedures reported in Chapter IV indicated that the subsamples chosen were not significantly different from the experimental population on the reading measure, the subsamples are not, in fact, exactly the same groups as those on which the reading analysis was made.

compositions written by the experimental and control groups. There were no statistically significant differences (p = .05) between the experimental and control groups on any of the nine comparisons made.

a. <u>main effects and interactions</u>. As Tables 17, 18, and 19 show, there were no significant differences in the writing quality of the experimental and control groups on the pre-mid, mid-post, or pre-post comparisons either when the grade-nine and -ten groups were examined separately or when the two grades were combined. The very high probability figures, ranging from .21 to .97 with the majority above .50, suggest that differences between the experimental and control groups on this measure were indeed small. Only one of the eighteen interactions examined was statistically significant at the .05 level of confidence. On the pre-post comparison there was a statistically significant interaction between the treatment and the sex of the subject on the grade-ten level. This must be regarded as spurious in light of the large number of interactions tested.

b. <u>changes in quality scores</u>. The changes in writing quality during the experiment, as Tables 17, 18, and 19 show, favor the control groups on eight of the nine comparisons made. These differences, however, are generally very small and statistically non-significant. Table 19 shows that over the six months of the experiment, the experimental students showed losses of almost one-half point (or four percent on

TABLE 17

Pre-mid comparison of writing quality measure showing the statistical probability for the treatment and for the teacher and sex interactions. Changes on the quality measure over the first three months of the experiment are given by grade, treatment, and sex.

	Grade 9		Grade 10		Grades Combined	
Statistical Data*	n	prob	n	prob	n	prob
Treatment	36	0.57	36	0.60	72	0.97
Teacher Interaction	36	0.70	36	0.31	72	0.48
Sex Interaction	36	0.86	36	0.28	72	0.64
Reading	103	0.16	91	0.004**	194	0.002**
Writing Change						
Experimental	n	raw score	n	raw score	n	raw score
Boys	8	-0.21	13	-0.28	21	-0.25
Girls	10	+0.52	5	-0.11	15	+0.31
Combined	18	+0.20	18	-0.24	36	-0.02
Control						
Boys	6	+0.13	6	-0.49	12	-0.17
Girls	12	+0.74	12	+0.38	24	+0.56
Combined	18	+0.54	18	+0.09	36	+0.31

*Only the probabilities of the differences between the experimental and control groups are presented here. The degrees of freedom and F-ratios are presented in Appendix B with the discussion of the null hypotheses.

**$p < .05$

the eleven-point scale) while the control students showed only very slight losses. Other differences between the experimental and control groups range from .01 to .45 points,

TABLE 18

Mid-post comparison of the writing quality scores showing the statistical probability for the treatment and for the teacher and sex interactions. Changes on the quality measure over the second three months of the experiment are given by grade, treatment, and sex.

	Grade 9		Grade 10		Grades Combined	
Statistical Data	n	prob	n	prob	n	prob
Treatment	36	0.74	36	0.69	72	0.96
Teacher Interaction	36	0.58	36	0.32	72	0.67
Sex Interaction	36	0.28	36	0.46	72	0.91
Reading	103	0.11	91	0.27	194	0.68
Writing Change						
Experimental	n	raw score	n	raw score	n	raw score
Boys	8	-0.60	13	-0.36	21	-0.45
Girls	10	-0.26	5	-0.34	15	-0.28
Combined	18	-0.41	18	-0.35	36	-0.38
Control						
Boys	6	+1.02	6	-1.44	12	-0.21
Girls	12	-0.94	12	+0.04	24	-0.45
Combined	18	-0.29	18	-0.45	36	-0.37

all very small differences.

When the boys and girls in the study were examined separately, only random variation of the scores was observed, and no consistent relationship between the treatment and changes on the quality measure was evident.

c. <u>discussion</u>. The hypothesized gains in writing

TABLE 19

Pre-post comparison of the writing quality measure showing the statistical probability for the treatment and for the teacher and sex interactions. Changes on the quality measure over the full six months of the experiment are given by grade, treatment, and sex.

	Grade 9		Grade 10		Grades Combined	
Statistical Data	n	prob	n	prob	n	prob
Treatment	36	0.45	36	0.21	72	0.86
Teacher Interaction	36	0.53	36	0.99	72	0.47
Sex Interaction	36	0.33	36	0.02*	72	0.94
Reading	103	0.02*	96	0.02*	194	0.001*
Writing Change						
Experimental	n	raw score	n	raw score	n	raw score
Boys	8	-0.81	13	-0.64	21	-0.70
Girls	10	+0.26	5	-0.45	15	+0.02
Combined	18	-0.21	18	-0.59	36	-0.40
Control						
Boys	6	+1.15	6	-1.93	12	-0.39
Girls	12	-0.20	12	+0.41	24	+0.11
Combined	18	+0.25	18	-0.37	36	-0.06

*$p < .05$

quality for the treatment group in the experiment were not confirmed. Gains in writing quality, although not statistically significant, generally favored the control group. Nor did any pattern emerge on the pre-mid, mid-post and pre-post comparisons to suggest that a delayed improvement in composition quality could be expected. The posttests were

of lower quality than the midtests.

2. Syntactic Density Analysis. Nine of the thirty-six null hypotheses tested in the study examined differences in the syntactic density of the compositions written by the experimental and control groups. There were no statistically significant differences between the experimental and control groups on any of the nine comparisons made.

a. main effects and interactions. As Tables 20, 21, and 22 show, there were no significant differences in the syntactic density of the experimental and control groups on the pre-mid, mid-post, or pre-post comparisons, either when the grade-nine and -ten groups were examined separately or when the two grades were combined. The high probability figures, ranging from .10 to .96 with the majority above .40, suggest that differences between the experimental and control groups on this measure were indeed small.

There were no significant interactions between the treatment and teacher variable or between the treatment and sex variable for any grade grouping on any test comparison. The probability figures are again very high, ranging from .21 to .83, suggesting that neither the sex of the subject nor the teacher who taught one of the four pairs of experimental and control classes substantially influenced the findings for the syntactic density analysis.

b. changes in syntactic density. The changes in syntactic density during the experiment, as Tables 20, 21, and 22 show,

TABLE 20

Pre-mid comparison of the syntactic density measure showing the statistical probability for the treatment and for the teacher and sex interactions. Changes on the syntactic density measure over the first three months of the experiment are given by grade, treatment, and sex.

	Grade 9		Grade 10		Grades Combined	
Statistical Data	n	prob	n	prob	n	prob
Treatment	48	0.45	48	0.85	96	0.49
Teacher Interaction	48	0.32	48	0.32	96	0.56
Sex Interaction	48	0.75	48	0.35	96	0.48
Reading	103	0.16	91	0.004*	194	0.002*
Writing Change						
Experimental	n	raw score	n	raw score	n	raw score
Boys	11	+0.38	16	+0.09	27	+0.20
Girls	13	+0.06	8	+0.28	21	+0.15
Combined	24	+0.21	24	+0.15	48	+0.18
Control						
Boys	9	+0.24	9	-0.01	18	+0.12
Girls	15	-0.20	15	+0.66	30	+0.23
Combined	24	-0.04	24	+0.41	48	+0.19

*$p < .05$

favor the experimental groups on six of the nine comparisons made. These differences, however, are generally very small and statistically non-significant. As Table 22 shows, over the six months of the experiment the experimental students made very slight gains (only one-tenth

TABLE 21

Mid-post comparison of the syntactic density measure showing the statistical probability for the treatment and for the teacher and sex interactions. Changes on the syntactic density measure over the second three months of the experiment are given by grade, treatment, and sex.

	Grade 9		Grade 10		Grades Combined	
Statistical Data	n	prob	n	prob	n	prob
Treatment	48	0.96	48	0.10	96	0.18
Teacher Interaction	48	0.83	48	0.60	96	0.46
Sex Interaction	48	0.45	48	0.52	96	0.38
Reading	103	0.11	91	0.23	194	0.68
Writing Change						
Experimental	n	raw score	n	raw score	n	raw score
Boys	11	+0.12	16	+0.00	27	+0.05
Girls	13	-0.43	8	+0.23	21	-0.18
Combined	24	-0.18	24	+0.08	48	-0.05
Control						
Boys	9	-0.35	9	-0.76	18	-0.56
Girls	15	-0.06	15	-0.05	30	-0.05
Combined	24	-0.17	24	-0.32	48	-0.24

of a point[1]) while the control students showed slight losses. The largest difference recorded, .25 favoring the grade-ten control group on the pre-mid comparison, has a statistical probability of .85.

---

[1]As noted in Chapter II, .8 points represent roughly one year's growth on Golub's conversion scale.

TABLE 22

Pre-post comparison of the syntactic density measure showing the statistical probability for the treatment and for the teacher and sex interactions. Changes on the syntactic density measure over the six months of the experiment are given by grade, treatment, and sex.

	Grade 9		Grade 10		Grades Combined	
Statistical Data	n	prob	n	prob	n	prob
Treatment	48	0.41	48	0.42	96	0.26
Teacher Interaction	48	0.21	48	0.56	96	0.57
Sex Interaction	48	0.96	48	0.21	96	0.24
Reading	103	0.20*	91	0.20*	194	0.001*
Writing Change						
Experimental	n	raw score	n	raw score	n	raw score
Boys	11	+0.51	16	+0.09	27	+0.26
Girls	13	-0.37	8	+0.52	21	-0.03
Combined	24	+0.03	24	+0.23	48	+0.13
Control						
Boys	9	-0.11	9	-0.77	18	-0.44
Girls	15	-0.26	15	+0.62	30	+0.18
Combined	24	-0.20	24	+0.10	48	-0.05

*p < .05

When the boys and girls in the study were examined separately, only random variation of the scores was observed, and no consistent relationship between the treatment and changes on the syntactic density measure was evident.

c. discussion. The hypothesized gains in syntactic

density for the treatment group in the experiment were not confirmed. Gains in syntactic density favored the experimental groups on two-thirds of the comparisons, but all differences were small. Nor did any pattern emerge on the pre-mid, mid-post, and pre-post comparisons to suggest that delayed improvement on the syntactic density measure could be expected.

3. T-Unit Length Analysis. Nine of the thirty-six null hypotheses tested in the study examined changes in the average T-unit length of the compositions written by the experimental and control students. Unlike the quality and syntactic density analyses discussed above, where subsamples of the compositions gathered in the study were analyzed, the compositions of all 194 students taking part in the study were analyzed for T-unit length. Statistically significant differences between the experimental and control groups were found on only one of the nine comparisons made. This difference favored the control group.

a. main effects and interactions. As Tables 23, 24, and 25 show, only the grade-ten experimental and control groups on the pre-mid comparison were significantly different at the .05 level of confidence. This difference favored the control group. There were no other significant differences in the T-unit length of the experimental and control compositions on the pre-mid, mid-post, or pre-post comparisons, either when the grade-nine and -ten groups were examined separately or when the two grades were combined.

TABLE 23

Pre-mid comparison of T-unit length showing the statistical probability for the treatment and for the teacher and sex interactions. Changes in average T-unit length over the first three months of the experiment are given by grade, treatment, and sex.

	Grade 9		Grade 10		Grades Combined	
Statistical Data	n	prob	n	prob	n	prob
Treatment	103	0.43	91	0.04*	194	0.32
Teacher Interaction	103	0.99	91	0.08	194	0.06
Sex Interaction	103	0.78	91	0.008*	194	0.05*
Reading	103	0.16	91	0.004*	194	0.002*
Writing Change						
Experimental	n	raw score	n	raw score	n	raw score
Boys	28	+0.65	25	-0.28	53	+0.21
Girls	22	-0.19	19	-0.64	41	-0.40
Combined	50	+0.28	44	-0.44	94	-0.06
Control						
Boys	24	+0.28	23	-0.16	47	+0.07
Girls	29	-0.84	24	+2.09	53	+0.49
Combined	53	-0.33	47	+0.99	100	+0.29

$^*p < .05$

The high probability figures, ranging from .14 to .59 with the majority above .40, suggest that differences between the experimental and control groups on the T-unit measure were indeed small. Although there were no statistically significant interactions between the treatment and the teacher variables for any grade grouping on any test

TABLE 24

Mid-post comparison of T-unit length showing the statistical probability for the treatment and for the teacher and sex interactions. Changes in average T-unit length over the second three months of the experiment are given by grade, treatment, and sex.

	Grade 9		Grade 10		Grades Combined	
Statistical Data	n	prob	n	prob	n	prob
Treatment	103	0.56	91	0.15	194	0.39
Teacher Interaction	103	0.22	91	0.10	194	0.10
Sex Interaction	103	0.63	91	0.99	194	0.71
Reading	103	0.11	91	0.27	194	0.68
Writing Change						
Experimental	n	raw score	n	raw score	n	raw score
Boys	28	+0.31	25	-0.24	53	+0.05
Girls	22	+0.07	19	-0.14	41	-0.03
Combined	50	+0.20	44	-0.19	94	+0.02
Control						
Boys	24	-0.31	23	-0.87	47	-0.58
Girls	29	+1.43	24	-1.03	53	+0.32
Combined	53	+0.65	47	-0.95	100	-0.11

comparison, significant interactions between the treatment and the sex of the subject were observed on both the pre-mid and pre-post comparisons for both the grade-ten students separately and for the grade-nine and -ten students combined. Both the significant interactions and the significant treatment variable noted above result from the low pretest scores of the grade-ten control girls, as will be discussed

49

[illegible]

TABLE 25

Pre-post comparison of T-unit length showing the statistical probability for the treatment and for the teacher and sex interactions. Changes in average T-unit length over the six months of the experiment are given by grade, treatment, and sex.

	Grade 9		Grade 10		Grades Combined	
Statistical Data	n	prob	n	prob	n	prob
Treatment	103	0.60	91	0.14	194	0.49
Teacher Interaction	103	0.65	91	0.30	194	0.31
Sex Interaction	103	0.97	91	0.02*	194	0.04*
Reading	103	0.02*	91	0.02*	194	0.001*
Writing Change						
Experimental	n	raw score	n	raw score	n	raw score
Boys	28	+0.95	25	-0.52	53	+0.26
Girls	22	-0.12	19	-0.78	41	-0.43
Combined	50	+0.48	44	-0.63	94	-0.04
Control						
Boys	24	-0.03	23	-1.03	47	-0.51
Girls	29	+0.59	24	+1.07	53	+0.81
Combined	53	+0.31	47	+0.04	100	+0.19

$*p < .05$

below. Since these are isolated findings attributable to the low pretest scores of one subgroup, no definite relationship between the treatment and T-unit length can be established.

b. <u>changes in T-unit scores</u>. The changes in average T-unit length during the experiment, as Tables 23, 24, and

25 show, favor the control groups on five of the nine comparisons made. The differences between the two groups are generally less than .5 words per T-unit.[1] The grade-ten experimental girls, however, showed gains of 2.09 words per T-unit on the pre-mid comparison and 1.07 words per T-unit on the pre-post comparison. Examination of the pretest scores of the grade-ten control girls (Appendix B, Table 70) suggests that these gains can be attributed to regression towards the mean. On the pretest, this group wrote an average of 13.7 words per T-unit, two words per T-unit below the grade-ten mean. On the mid- and posttests, however, they wrote 15.8 and 14.7 words per T-unit respectively, both above the grade-ten mean. These gains of the grade-ten control girls, as noted above, account for both the significant treatment variable and the significant interactions between the treatment and the sex of the subject.

When the boys and girls are examined separately, other changes on the T-unit measure appear random, sometimes favoring the experimental groups and sometimes favoring the control groups, and no consistent relationship between the treatment and changes on the T-unit measure is evident.

c. discussion. The hypothesized gains in average T-unit length for the treatment group in the experiment were not confirmed. The only comparison to reach statistical

---

[1]As was noted in Chapter II, .7 words per T-unit appear to represent one year's growth on Hunt's (1965) norms.

significance favored the control group and appeared to be the result of regression toward the mean. Nor did any pattern emerge on the pre-mid, mid-post, and pre-post comparisons to suggest that a delayed increase in average T-unit length could be expected. The T-unit lengths in the compositions of the experimental group were slightly shorter on the posttest than on the pretest, while the control group showed slight gains over this period.

4. Fluency Analysis. Nine of the thirty-six null hypotheses tested in the study examined changes in the average length of the compositions written by the experimental and control groups. The compositions of all 194 students taking part in the study were analyzed on this measure. Statistically significant differences between the experimental and control groups were found on two of the nine comparisons made, both differences favoring the control group.

a. main effects and interactions. As Tables 26, 27, and 28 show, at the .05 level of confidence the treatment variable was statistically significant on only two of the nine comparisons, the mid-post comparison for both the grade-ten group separately and the grade-nine and -ten groups combined. Both of these favor the control group and are the result of substantial losses by the experimental group compared with moderate losses by the control group. The other probabilities range from .12 to .98 with the majority above .40, suggesting that differences between the experimental and

TABLE 26

Pre-mid comparison of the fluency measure showing the statistical probability for the treatment and for the teacher and sex interactions. Changes on the fluency measure over the first three months of the experiment are given by grade, treatment, and sex.

	Grade 9		Grade 10		Grades Combined	
Statistical Data	n	prob	n	prob	n	prob
Treatment	103	0.12	91	0.39	194	0.19
Teacher Interaction	103	0.18	91	0.99	194	0.71
Sex Interaction	103	0.93	91	0.35	194	0.36
Reading	103	0.16	91	0.004*	194	0.002*
Writing Change						
Experimental	n	raw score	n	raw score	n	raw score
Boys	28	+17.1	25	+22.0	53	+19.4
Girls	22	+14.6	19	+17.9	41	+16.1
Combined	50	+16.0	44	+20.2	94	+18.0
Control						
Boys	24	-6.8	23	-15.5	47	-11.1
Girls	29	-13.2	24	-0.4	53	-7.4
Combined	53	-10.3	47	-7.8	100	-9.1

*$p < .05$

control groups on the fluency measure were small.

While no significant interactions between the treatment and the teacher variable were found, a significant interaction between the treatment and the sex of the subject was found for the grade-ten group on both the mid-post and pre-post comparisons. Both the significant treatment

TABLE 27

Mid-post comparison of the fluency measure showing the statistical probability for the treatment and for the teacher and sex interactions. Changes on the fluency measure over the second three months of the experiment are given by grade, treatment, and sex.

	Grade 9		Grade 10		Grades Combined	
Statistical Data	n	prob	n	prob	n	prob
Treatment	103	0.40	91	0.04*	194	0.04*
Teacher Interaction	103	0.39	91	0.51	194	0.52
Sex Interaction	103	0.95	91	0.03*	194	0.18
Reading	103	0.11	91	0.26	194	0.68
Writing Change						
Experimental	n	raw score	n	raw score	n	raw score
Boys	28	-60.2	25	-31.3	53	-46.6
Girls	22	-33.1	19	-60.5	41	-45.8
Combined	50	-48.3	44	-43.9	94	-46.2
Control						
Boys	24	-41.0	23	-16.0	47	-28.8
Girls	29	-12.9	24	-4.0	53	-8.8
Combined	53	-25.6	47	-9.8	100	-18.2

*p $<$ .05

variable and the significant interactions are the result of the grade-ten experimental girls who showed considerable losses on the mid-post and pre-post comparisons while the grade-ten control girls showed only moderate losses. In light of the large number of comparisons made, the apparent significance of these results must be regarded as spurious.

TABLE 28

Pre-post comparison of the fluency scores showing the statistical probability for the treatment and for the teacher and sex interactions. Changes on the fluency measure over the six months of the experiment are given by grade, treatment, and sex.

	Grade 9		Grade 10		Grades Combined	
Statistical Data	n	prob	n	prob	n	prob
Treatment	103	0.42	91	0.72	194	0.99
Teacher Interaction	103	0.92	91	0.58	194	0.69
Sex Interaction	103	0.96	91	0.04*	194	0.14
Reading	103	0.02*	91	0.02*	194	0.001*
Writing Change						
Experimental	n	raw score	n	raw score	n	raw score
Boys	28	-43.1	25	-9.3	53	-27.2
Girls	22	-18.5	19	-42.6	41	-29.7
Combined	50	-32.3	44	-23.7	94	-28.3
Control						
Boys	24	-47.8	23	-31.5	47	-39.8
Girls	29	-26.0	24	-4.4	53	-16.2
Combined	53	-35.9	47	-17.7	100	-27.3

*$p < .05$

b. <u>changes in fluency scores</u>. On the pre-mid comparison, a relationship between the treatment and the fluency scores appeared to develop. The experimental students, both boys and girls on both grade levels, showed gains on the fluency measure, while the control students experienced losses. Overall, the gains were 18 words per composition for the

experimental students, and losses were an average of nine words per composition for the control students, as Table 26 shows. On the mid-post comparison, however, as Table 27 shows, both the experimental and control groups showed losses, the experimental groups over 46 words per composition but the control group only 18 words per composition, making the experimental and control losses over the six months of the experiment, as Table 28 shows, almost exactly the same. The very large losses for the experimental groups on the mid-post comparison were statistically significant at the .05 level of confidence but appear to be merely the result of losing on the mid-post comparison what they had gained on the pre-mid comparison.

When the boys and the girls in the study were examined separately, no consistent relationship between the treatment and the fluency measure was evident. The average change scores vary widely, ranging from gains of 22 words per composition to losses of 60 words per composition. Only the grade-ten control girls showed consistent averages, with average losses of four words per composition or fewer on each of the three test comparisons.

c. discussion. The hypothesized gains in fluency for the treatment group in the experiment were not confirmed. While changes in fluency on the pre-mid comparison favored the experimental group, changes on both the mid-post and pre-post comparisons favored the control group. The experimental and control groups experienced almost equal

losses on the fluency measure over the six months of the experiment.

5. <u>Analysis of the Writing Measures Using Three Levels of Student Ability</u>. To see if the treatment had differing effects for the high-, medium-, and low-ability students in the sample, two-way analysis of variance was used to compare the experimental and control students on the reading and the four writing measures.[1]

Reference to Table 29, the analysis of the grade-nine groups, and Table 30, the analysis of the grade-ten groups, shows that there was no statistically significant interaction between the treatment and ability level for any of the five measures on any of the three test comparisons. The highest statistical probability reached on any variable was .16 for the grade-nine students on the mid-post fluency comparison. Other probabilities generally ranged between .4 and .7, suggesting that the effects of the treatment were not significantly influenced by the ability of the subject.

B. <u>Analysis of Major Sentence Errors</u>

A sample of the compositions collected in the study (the compositions of the 36 experimental and 36 control

---

[1]The subjects were divided on the basis of ability on the first test in each comparison (pre-mid, mid-post, and pre-post). The effects of the treatment were then tested for each ability level. The results are expressed as the statistical significance of the interaction between the treatment and the ability level.

TABLE 29

Two-way analysis of variance with grade-nine subjects divided into high-, medium-, and low-ability on the reading measure and the four writing measures

Measure	Criterion Test	Classification Test	n	D.F.	F-Ratio	Probability	Decision
Quality	Mid	Pre	36	2	0.569	0.572	N.S.
	Post	Pre	36	2	0.813	0.453	N.S.
	Post	Mid	36	2	0.393	0.678	N.S.
Syntactic Density	Mid	Pre	48	2	1.499	0.235	N.S.
	Post	Pre	48	2	0.498	0.611	N.S.
	Post	Mid	48	2	0.803	0.455	N.S.
T-Unit Length	Mid	Pre	103	2	0.688	0.505	N.S.
	Post	Pre	103	2	0.185	0.831	N.S.
	Post	Mid	103	2	1.093	0.339	N.S.
Fluency	Mid	Pre	103	2	1.072	0.346	N.S.
	Post	Pre	103	2	1.101	0.337	N.S.
	Post	Mid	103	2	1.890	0.156	N.S.
Reading	Mid	Pre	103	2	1.589	0.209	N.S.
	Post	Pre	103	2	1.092	0.340	N.S.
	Post	Mid	103	2	0.208	0.812	N.S.

TABLE 30

Two-way analysis of variance with grade-ten subjects divided into high-, medium-, and low-ability on the reading and the four writing measures

Measure	Criterion Test	Classification Test	n	D.F.	F-Ratio	Probability	Decision
Quality	Mid	Pre	36	2	0.549	0.583	N.S.
	Post	Pre	36	2	0.197	0.823	N.S.
	Post	Mid	36	2	0.775	0.470	N.S.
Syntactic Density	Mid	Pre	48	2	0.374	0.690	N.S.
	Post	Pre	48	2	0.698	0.503	N.S.
	Post	Mid	48	2	0.474	0.626	N.S.
T-Unit Length	Mid	Pre	91	2	0.919	0.403	N.S.
	Post	Pre	91	2	0.196	0.822	N.S.
	Post	Mid	91	2	0.546	0.581	N.S.
Fluency	Mid	Pre	91	2	0.831	0.439	N.S.
	Post	Pre	91	2	0.426	0.655	N.S.
	Post	Mid	91	2	0.979	0.380	N.S.
Reading	Mid	Pre	91	2	0.290	0.749	N.S.
	Post	Pre	91	2	0.505	0.605	N.S.
	Post	Mid	91	2	1.627	0.203	N.S.

students used in the quality analysis) was analyzed for major sentence errors: run-on sentences, sentence fragments, errors in subject-verb agreement, and errors in pronoun agreement. These errors were counted for each paper and calculated for each hundred words of text. Following adjustment for topic difficulty (see Appendix F), the error-reduction scores for the experimental and control groups were compared for the pre-mid, mid-post, and pre-post comparisons by means of t-tests.

As Table 31 shows, at the .05 level of confidence, differences in the error-reduction scores of the experimental and control groups were statistically significant on only one of the nine comparisons made. This difference favored the grade-nine and -ten experimental groups combined on the mid-post comparison.

Examination of Table 31 shows that all error reduction on the pre-mid comparison was slight with probabilities ranging from .60 to .80, all in favor of the control groups. Error reduction on both the mid-post and pre-post comparisons, however, favored the experimental groups and was considerably more significant with the majority of probabilities .20 and greater. Only one of these comparisons, however, reached the .05 level of confidence.

As Table 32 shows, the experimental groups combined made more errors per hundred words on both the pre- and midtests but fewer errors on the posttest than the control groups did. While the experimental groups showed considerable

TABLE 31

Major error analysis by grade and test period with scores adjusted for topic difficulty

Test	n	Experimental $\bar{x}$	Experimental σ	Control $\bar{x}$	Control σ	Experimental Minus Control	t	df	Decision	Prob
Grades 9 & 10 Combined										
Pre-Mid	36	0.1294	1.4774	-0.0269	1.0433	0.1563	0.521	70	N.S.	.70
Mid-Post	36	-0.5376	1.3883	0.0890	0.9031	-0.6266	-2.206	70	Sig.	.05*
Pre-Post	36	-0.4082	0.7939	0.0621	1.1634	-0.4703	-1.962	70	N.S.	.10
Grade 10										
Pre-Mid	18	0.3004	1.9195	0.1669	1.0401	0.1335	0.263	34	N.S.	.80
Mid-Post	18	-0.7728	1.7483	-0.0300	0.9532	-0.7428	-1.550	34	N.S.	.20
Pre-Post	18	-0.4725	0.9212	0.1368	1.1838	-0.6093	-1.677	34	N.S.	.20
Grade 9										
Pre-Mid	18	-0.0415	0.8647	-0.2206	1.0391	0.1791	0.567	34	N.S.	.60
Mid-Post	18	-0.3024	0.8914	0.2080	0.8605	-0.5104	-1.699	34	N.S.	.10
Pre-Post	18	-0.3439	0.6635	-0.0126	1.1719	-0.3313	-1.024	34	N.S.	.40

*p < .05

TABLE 32

Major errors per hundred words by treatment and grade with scores adjusted for topic difficulty.

	n	Pretest $\bar{x}$	Pretest σ	Midtest $\bar{x}$	Midtest σ	Posttest $\bar{x}$	Posttest σ
Experimental							
Grade 9	18	0.95	0.59	0.90	0.54	0.60	0.64
Grade 10	18	1.45	1.40	1.76	1.75	0.99	1.06
9 & 10 Combined	36	1.20	1.09	1.33	1.35	0.79	0.88
Control							
Grade 9	18	1.19	0.88	0.97	0.72	1.18	0.77
Grade 10	18	0.83	0.81	1.00	1.12	0.97	1.19
9 & 10 Combined	36	1.01	0.85	0.98	0.93	1.07	0.99

fluctuations over the three test periods, the control groups maintained a constant error ratio. When the grade-nine groups were examined separately, the experimental students both made fewer errors at each test period and showed more consistent error reduction than the control students did. On the grade-ten level, on the other hand, the control students made fewer errors than the experimental students did at each test period and also were more consistent than the experimental students were over the three test periods. The treatment, then, did not appear to influence major sentence errors in the compositions of the experimental students. Although the experimental students showed greater reduction in these errors than the control students did, the changes were neither consistent when the grade-nine and -ten groups were examined separately nor statistically significant on eight

of the nine comparisons made.

While no further inference is warranted, beyond that just stated, these experimental-control comparisons come closer than any others in the study to support for the hypothesized but uncorroborated influence of reading skill on writing skill.

C. Relationships Between the Reading and Writing Measures

Two methods were used in the current study to examine the relationships between the reading and writing measures. First, correlations between the reading and writing measures were computed with the sample divided by grade, treatment, and sex. Second, the average scores of each subgroup on the reading and the four writing measures were compared.

Correlations between the reading and quality measures and the reading and fluency measures for the entire sample were significant beyond the .001 level of confidence,[1] confirming previous correlational data reported in Chapter II.

---

[1]The statistical significance of a correlation coefficient is dependent upon the number of subjects correlated. The sample sizes for the correlations reported in this chapter vary from 33 to 194 and therefore it might be misleading to compare the statistical probabilities on such measures as the quality measure (n = 33 for the boys in the study) and the fluency measure (n = 100 for the boys in the study). Consequently, the correlation coefficients and not the probabilities are used when measures are compared.

Selected correlations are presented in this section for the total sample, the boys and the girls, and the experimental and control groups. The complete correlational data obtained in the study, a total of 324 correlation coefficients, is presented in Appendix I.

However, the correlations between the reading and syntactic density measures and the reading and T-unit measures were very small and did not approach statistical significance.

When the average scores of the experimental and control groups on the reading and writing measures were compared, no consistent pattern emerged: the group with the highest average scores on the reading measure was generally not the group with the highest average scores on any of the writing measures.

1. Correlations Between Reading and Writing Measures. When the total sample was examined, correlations between the reading and quality measures were highly significant, higher for the girls than for the boys, and higher for the control group than for the experimental group. Correlations between the reading and fluency measures were not so high as those between the reading and quality measures, and were generally higher for the boys than for the girls in the study. Correlations between the reading measure and the syntactic density and T-unit measures, on the other hand, were very low. As might be expected from the high correlations between the syntactic density and T-unit measures reported in Chapter IV, the correlation between the reading and each of these writing measures was substantially the same.

a. correlations for the total sample. As Table 33 shows, the highest correlations observed for the total sample were between the reading and quality measures. The correlation

TABLE 33

Correlations between the reading and the four writing measures on the pre-, mid-, and posttests for the total experimental population

Writing Measures	Reading Measures								
	Pretest			Midtest			Posttest		
	n	r	prob	n	r	prob	n	r	prob
Quality	72	.47	.0001	72	.54	.0001	72	.48	.0001
Syntactic Density	96	.02	.82	96	.07	.51	96	.09	.34
T-Unit	194	.06	.43	194	.08	.28	194	.09	.19
Fluency	194	.20	.004	194	.19	.008	194	.27	.0001

coefficients, ranging from .47 to .54, were significant beyond the .001 level of confidence for all three test periods. The fluency measure also correlated significantly ($p < .008$) with the reading measure, but the correlation coefficients, ranging from .19 to .27, were not so high as they were for the reading and quality measures. The correlations between the reading and syntactic density measures and the reading and T-unit measures, on the other hand, were all very small, none with a correlation coefficient reaching .10.

b. <u>correlations for the boys and girls separately</u>. Comparison of Tables 34 and 35 shows that correlations between the reading and quality measures were much higher for the girls in the study than for the boys. Correlation coefficients for the girls ranged from .57 to .70 ($p < .0001$), while the correlation coefficients for the boys ranged from .33 ($p > .05$) to .37 ($p < .05$). This confirms earlier data reported by Fishco (1966) who found that correlations for the girls in his study were significant at the .01 level of confidence, but the correlations for the boys did not reach the .05 level of confidence.

The correlations between the reading and fluency measures, however, were higher for the boys than for the girls on both the pre- and posttests. On both these tests the correlations were significant beyond the .01 level of confidence for the boys but not statistically significant at the .05 level of confidence for the girls. The correlations

TABLE 34

Correlations between the reading and the four writing measures on the pre-, mid-, and posttests for the boys in the study

Writing Measures	Reading Measures								
	Pretest			Midtest			Posttest		
	n	r	prob	n	r	prob	n	r	prob
Quality	33	.35	.05	33	.33	.06	33	.37	.03
Syntactic Density	45	.13	.39	45	.22	.15	45	.12	.44
T-Unit	100	.12	.24	100	.12	.24	100	.15	.14
Fluency	100	.26	.01	100	.15	.13	100	.36	.0001

TABLE 35

Correlations between the reading and the four writing measures on the pre-, mid-, and posttests for the girls in the study

Writing Measures	Reading Measures								
	Pretest			Midtest			Posttest		
	n	r	prob	n	r	prob	n	r	prob
Quality	39	.57	.0001	39	.70	.0001	39	.59	.0001
Syntactic Density	51	-.06	.66	51	-.05	.73	51	.09	.54
T-Unit	94	-.01	.93	94	.04	.69	94	.04	.72
Fluency	94	.15	.14	94	.22	.03	94	.17	.10

between the reading and syntactic-density measures and the reading and T-unit measures were also higher for the boys than for the girls, but none of these correlations approached significance at the .05 level of confidence.

c. correlations for the experimental and control groups separately. Comparison of Tables 36 and 37 indicates that correlations between the reading and quality measures were much higher for the control groups than for the experimental groups at all three test periods. The correlation coefficients on these two measures ranged from .56 to .72 for the control groups but from .23 to .36 for the experimental groups. The correlations between the reading and fluency measures, on the other hand, were higher for the experimental groups than they were for the control groups on both the mid- and posttests.[1]

2. Relationships Between Reading and Writing Measures: Group Averages and Change Scores. When the treatment effects were examined, as noted above, statistically significant changes

---

[1]The investigator can offer no explanation for this as an effect of the treatment. It might be suggested that some of the experimental students improved on the reading measure but not on the quality measure on the mid- and posttests which would tend to introduce more variation to the matrix of scores and thus lower the correlation coefficient. This does not, however, account for the increase in the correlation coefficients for the control groups, nor does it account for the fact that the control groups had substantially higher correlation coefficients on the pretest than the experimental students did. And furthermore, it does not account for the increase in correlation coefficients for the experimental group on the mid- and post-fluency measure.

TABLE 36

Correlations between the reading and the four writing measures on the pre-, mid-, and posttests for the experimental students

Writing Measures	Reading Measures								
	Pretest			Midtest			Posttest		
	n	r	prob	n	r	prob	n	r	prob
Quality	36	.36	.03	36	.23	.19	36	.30	.08
Syntactic Density	48	.13	.40	48	.03	.86	48	.20	.17
T-Unit	94	.16	.13	94	-.02	.82	94	.14	.20
Fluency	94	.19	.06	94	.34	.001	94	.33	.001

TABLE 37

Correlations between the reading and the four writing measures on the pre-, mid-, and posttests for the control students

Writing Measures	Reading Measures								
	Pretest			Midtest			Posttest		
	n	r	prob	n	r	prob	n	r	prob
Quality	36	.56	.0001	36	.72	.0001	36	.60	.0001
Syntactic Density	48	-.10	.52	48	.08	.57	48	.02	.91
T-Unit	100	-.04	.69	100	.18	.08	100	.05	.60
Fluency	100	.21	.04	100	.06	.51	100	.21	.04

on the reading measure were not accompanied by statistically significant changes on the writing measures.

Comparison of the gains by the experimental and control groups on each measure reveals that in fewer than 50 percent of the cases, the group showing the largest gains on the reading measure also shows the largest gains on the writing measure. Furthermore, when the boys and the girls in the experimental and control groups are ranked on the reading and the four writing measures, a group's rank on the reading measure can be used to predict its rank on the writing measure in just over 28 percent of the cases, very close to the 25 percent that would be expected by chance alone, there being four groups.

Table 38 outlines the gains made by the experimental and control groups on the reading and the four writing measures on the pre-mid, mid-post, and pre-post comparisons. The group showing the largest gains on the reading measure also showed the largest gains on the writing measure on only sixteen of the thirty-six comparisons.

On the pretest,[1] when the grade-ten experimental and control boys and girls are ranked from one to four on the reading and each of the four writing measures, as Table 39 shows, a group's rank on the reading measure can be used to predict its rank on the writing measure on only four of the

---

[1]The pretest, of course, is unaltered by treatment. Therefore, one might expect any natural relationships between reading and writing to be most obvious here.

TABLE 38

Comparison of the gains shown by the experimental and control groups on the reading and the four writing measures

Grade	Measure	Pre-mid Reading	Pre-mid Writing	Mid-post Reading	Mid-post Writing	Pre-post Reading	Pre-post Writing
9 & 10	Quality	exp*	con	con	con	exp*	con
	S.D.	exp*	con	con	exp	exp*	exp
	T-Unit	exp*	con	exp	exp	exp*	con
	Fluency	exp*	exp	exp	con*	exp*	exp
9	Quality	exp	con	con	con	exp*	con
	S.D.	exp	exp	con	con	exp*	exp
	T-Unit	exp	exp	exp	con	exp*	exp
	Fluency	exp	exp	exp	con	exp*	exp
10	Quality	exp*	con	con	exp	exp*	con
	S.D.	exp*	con	con	exp	exp*	exp
	T-Unit	exp*	con*	con	exp	exp*	con
	Fluency	exp*	exp	con	con*	exp*	con

*Indicates group gains significant at .05 level of confidence. As was noted in Chapter IV, separate computations on the reading measure were not made for the subsamples used in the quality and syntactic density analyses. However, as was also noted in Chapter IV, the reading scores of these subsamples were not significantly different from the reading scores of the total experimental population.

TABLE 39

Grade Ten: Ranks of the experimental and control boys and girls on the reading measure and the four writing measures

		Quality		Syntactic Density		T-Unit		Fluency	
		Writing	Reading*	Writing	Reading	Writing	Reading	Writing	Reading
Experimental	Boys	4	3	2	3	3	4	3	4
	Girls	1	4	4	4	2	2	2	2
Control	Boys	2	2	1	2	1	3	1	3
	Girls	3	1	3	1	4	1	4	1

*Sample sizes for the quality and syntactic density measures differ from those for the T-unit and fluency measures. Consequently, in some cases the ranks on the reading measures differ.

sixteen comparisons. On the other hand, on three of the sixteen comparisons, the group ranking highest on one measure ranks lowest on the other. While in some cases there were only small differences in actual scores between two rankings, in other cases the differences were substantial. The experimental girls, for example, scored almost one standard deviation below the control girls on the initial reading measure, yet one-half standard deviation above the control girls on the initial quality measure. Nor does this appear to be an anomaly of testing, as the experimental girls maintained the high-quality and low-reading scores on both the mid- and posttests.[1]

As Table 40 shows, the rankings for the grade-nines are similar to those for the grade-tens: a group's rank on one measure can be used to predict its rank on the other measure on five of the sixteen comparisons. Again, some of the differences on the reading and writing measures are substantial. The experimental boys, for example, scored one-half standard deviation below the grade-nine mean on the initial reading measure yet one-half standard deviation above the mean on the quality measure, and maintained the below-average-reading and above-average-writing-quality scores on the mid- and posttests.

---

[1]The means and standard deviations for the experimental and control boys and girls on the reading and each of the four writing measures are presented in Appendix K.

TABLE 40

Grade Nine: Ranks of the experimental and control boys and girls on the reading measure and the four writing measures

		Quality		Syntactic Density		T-Unit		Fluency	
		Writing	Reading*	Writing	Reading	Writing	Reading	Writing	Reading
Experimental	Boys	1	4	4	4	4	4	4	4
	Girls	4	3	2	3	2	3	3	3
Control	Boys	2	1	3	2	3	1	2	1
	Girls	3	2	1	1	1	2	1	2

*Sample sizes for the quality and syntactic density measures differ from those for the T-unit and fluency measures. Consequently, in some cases the ránks on the reading measure differ.

D. Summary

No evidence was found in the study to support the hypothesis that growth in reading skill would be accompanied by growth in writing skill. Only three of the thirty-six null hypotheses tested in the study were rejected at the .05 level of confidence, approximately the number that would be expected to be rejected by chance alone. Furthermore, all three of the significant differences favored the control group.

Nor did a pattern develop on any of the four writing measures to suggest a relationship between the treatment and skill in composition. Gains on 20 of the test comparisons favored the control group while gains on the other 16 comparisons favored the experimental group. A pattern appeared to develop on the pre-mid fluency comparison where both the boys and girls in the grade-nine and -ten experimental groups showed gains while their control counterparts showed losses. However, on the mid-post comparison the experimental groups showed heavy losses and the control groups only moderate losses. Over the six months of the experiment, then, the two groups showed approximately equal losses.

Neither the sex of the subject nor the teacher who taught one of the four pairs of experimental and control groups in the study appeared to influence the results appreciably. At the .05 level of confidence, there were no significant interactions between the treatment and teacher

variables. Five of the thirty-six interactions between the treatment and the sex of the subject were significant at the .05 level of confidence; however, three of these were attributed to the grade-ten control girls who scored below the grade-ten mean on the pretest yet above the mean on both the mid- and posttests. Consequently, these interactions do not appear to have been influenced by the treatment.

Nor did the ability level of the student appear to influence the results on either the reading or the writing measures. Two-way analysis of variance showed no statistically significant interaction between the treatment and the ability level of the subject.

Only one of the nine comparisons on the major error measure revealed a statistically significant difference ($p = .05$) between the experimental and control groups. Changes on this measure, however, favored the experimental group on both the mid-post and pre-post comparisons, the only writing-measure comparisons in the study to yield results in the hypothesized direction.

Statistically significant correlations ($p = .05$) were observed between the reading and quality and reading and fluency measures, but the correlations between the reading and syntactic density and reading and T-unit measures were low and statistically non-significant. The correlations were generally higher for the girls than the boys in the study and for the control groups than for the experimental groups, but neither of these appear to have been influenced

by the treatment.

When the scores of the boys and girls in the experimental and control groups were ranked on the reading and each of the four major writing measures, there was no apparent relationship between the two sets of scores: the group scoring the highest on the reading measure seldom scored the highest on the writing measure.

CHAPTER VI

SUMMARY, CONCLUSIONS, AND IMPLICATIONS

A. Summary

This study examined the relationship between measured changes in reading ability and measured changes in writing ability at two grade levels over a six-month period. The experimental design called for two experimental and two control classes at each of the grade-nine and grade-ten levels. Following the administration of the initial reading test and collection of initial writing samples, the experimental students were administered the reading treatment. Parallel reading tests were given and parallel writing samples were taken three months and six months following the administration of the treatment. Four measures were used to evaluate the compositions: raters evaluated a 37-percent sample of the compositions for overall quality, a 50-percent sample was analyzed using the Syntactic Density Score, and all the compositions gathered in the study were analyzed for T-unit length and fluency. In addition the 37-percent sample used in the quality analysis was analyzed for major sentence errors. Since, despite efforts to select pairs of classes judged to be of comparable ability, and the random assignment of treatment to one class in each pair, the experimental and control groups were not in fact of equal ability at the onset of the experiment, as evidenced by both the initial reading

and writing scores, the statistical procedure of analysis of covariance was used to compare the experimental and control groups on the reading measure and the four writing measures. Interactions between the treatment and the sex of the subject and the treatment and the teacher who taught one of the four pairs of experimental and control groups were also examined. To test for effects at varying levels of ability, the experimental and control groups were divided into high-, medium-, and low-ability levels using the initial reading and writing measures, and the groups were compared using two-way analysis of variance. The experimental and control groups were compared on the reading and each of the four writing measures using three time periods: the initial three months (the pre-mid comparison), the second three months (the mid-post comparison), and the full six months of the experiment (the pre-post comparison). Finally, relationships between the reading and writing scores were examined both by computing correlations and by comparing the average scores on the reading and writing measures obtained when the groups were examined by treatment, grade, and sex.

1. Reading Changes. The current study examined changes in writing skill in the light of changes in reading skill. The changes in reading skill were obtained as part of a concurrent experiment (Martin and Belanger, 1977) and were reported in the current study only because they provided a basis for expecting changes in writing ability. The differences in the reading gains for the treatment and

non-treatment groups were as follows:

a. the grade-nine sample: Over the six months of the experiment the grade-nine experimental students showed reading gains superior to those of the control students statistically significant at the .020 level of confidence. The gains at the end of three months and between three and six months also favored the experimental groups, but neither achieved statistical significance at the .05 level of confidence. On the grade norms provided by Davis (1962), the 2.49 point superiority of the experimental students represented .41 of a year's growth over the six-month period.

b. the grade-ten sample: Over the first three months of the experiment and over the entire six months of the experiment the grade-ten experimental students showed reading gains superior to those of the control students at the .004 and the .023 level of confidence respectively. Slight and statistically non-significant differences on the three- to six-month comparison favored the control groups. On the grade norms provided by Davis (1962) the 3.40-point superiority of the experimental students represented .57 of a year's growth over the six-month period.

When the grade-nine and -ten samples were combined, the experimental groups were superior to the control groups at

the .002 level of confidence on the first three-month comparison and at the .001 level of confidence for the entire six-month comparison. Slight and statistically non-significant differences on the three- to six-month comparison also favored the combined experimental groups.

On the reading measure, statistical analysis revealed no statistically significant interactions between the treatment and the sex of the subject or the treatment and the teacher who taught one of the four pairs of experimental and control classes. Nor did the treatment have a differing effect on the high-, medium-, and low-ability students. Two-way analysis of variance revealed no significant differences when the experimental and control groups were compared using the three levels of ability.

2. Writing Changes. When the treatment and non-treatment groups were compared on the four major writing measures used in the study, only random comparisons reached statistical significance at the .05 level of confidence. Only three of the thirty-six null hypotheses examined in the study showed statistically significant differences in the writing of the experimental and control groups. All three of these differences favored the control groups. A subsidiary analysis of the reduction of major sentence errors, an analysis not expected to reveal significant differences between the experimental and control groups, was, in fact, the only writing measure used in the study which revealed differences favoring the experimental groups, but these differences were

generally not statistically significant at the .05 level of confidence.

a. <u>The Writing Quality Measure</u>: At the .05 level of confidence there were no statistically significant differences in the writing quality of the experimental and control groups on any of the three test comparisons, either when the grade-nine and grade-ten groups were examined separately or when the two grades were combined. Generally slight and statistically non-significant differences favored the control groups on eight of the nine null hypotheses tested.
One statistically significant interaction between the treatment and the sex of the subject was found. The control girls made modest gains over the six months of the experiment while the control boys experienced substantial losses, resulting in a treatment-sex interaction on the pre-post comparison significant at the .05 level of confidence.

b. <u>The Syntactic Density Measure</u>: At the .05 level of confidence, there were no statistically significant differences in the syntactic density of the papers written by the experimental and control students on any of the three test comparisons, either when the grade-nine and grade-ten groups were examined separately or when the two grades were combined. Generally slight and statistically non-significant differences favored the experimental groups on six

of the nine null hypotheses tested. No statistically significant interactions were observed between the treatment and the sex of the subjects at the .05 level of confidence.

c. <u>The T-unit Length Measure</u>: Only one of the nine null hypotheses on changes in T-unit length examined in the study reached statistical significance at the .05 level of confidence. This difference favored the grade-ten control group and was attributed to the grade-ten control girls' scoring below the grade-ten mean on the pretest yet above the mean on the midtest. Of the remaining eight null hypotheses tested, generally slight and statistically non-significant differences favored the experimental groups on four comparisons and the control groups on four comparisons. Below-average scores of the control girls on the pretest combined with above-average scores for this group on both the mid- and posttests led to four statistically significant but apparently spurious interactions between the treatment and the sex of the subject. These interactions were observed on both the pre-mid and pre-post comparisons for both the grade-ten groups and the grade-nine and -ten groups combined.

d. <u>The Fluency Measure</u>: Only two of the nine null hypotheses on changes in fluency tested in the study reached statistical significance at the .05

level of confidence. Both of these differences favored the control group. Although the experimental and control groups showed equal losses on the fluency measure over the six months of the experiment, wide fluctuations of the experimental group on the mid- and posttests resulted in significant differences between the experimental and control groups on the mid-post comparison. Statistically non-significant differences favored the experimental group on four of the remaining seven null hypotheses.

Two statistically significant interactions between the treatment and the sex of the subject were observed, both resulting from the modest but consistent losses of the control girls when compared with the wide fluctuations of the experimental girls on the fluency measure.

No statistically significant interactions were found between the treatment and the teacher who taught one of the four pairs of experimental and control classes in the study. Nor did the ability of the subjects influence the findings. When the subjects were divided into high-, medium-, and low-ability, no statistically significant interaction between the treatment and the ability of the subject was observed on any of the 36 null hypotheses tested.

Only one of the nine comparisons on the reduction of major sentence errors, the combined grade-nine and -ten classes on the mid-post comparison, revealed a statistically

significant difference between the experimental and control students at the .05 level of confidence. This difference favored the experimental students. All of the pre-mid gains in error reduction were slight and favored the control groups on both the grade-nine and grade-ten levels. On the mid-post and pre-post comparisons, on the other hand, the changes were more substantial and favored the experimental groups. The grade-nine and grade-ten students showed similar gains here. Although these changes were generally not statistically significant, they were in the hypothesized direction.

This study also proposed to examine the temporal relationship between changes in reading scores and changes in writing scores. The three test comparisons were designed to determine if the writing changes followed the reading changes promptly (the pre-mid comparison), followed the reading changes after some delay (the mid-post comparison), or accompanied the reading changes throughout the experiment (the pre-post comparison). Since the hypothesized writing gains were not found, this question could not be examined.

3. Relationships Between the Reading and Writing Scores. Relationships between the reading and writing scores obtained in the study were examined both in terms of traditional Pearson product-moment correlations and in terms of the average performance on each measure when the groups were divided by treatment, sex, and grade level.

Correlational data gathered in the current study generally supports that reported by previous researchers.

Correlations between the reading and quality measures for the entire sample were above .47, significant beyond the .0001 level of confidence. Correlations between the reading and fluency measures ranged from .19 to .27, less substantial than those for the reading and quality measures, yet significant beyond the .008 level of confidence. On the other hand, correlations between the reading and syntactic density measures and the reading and T-unit length measures were all very slight and often negative.

When the boys and the girls were compared separately, the correlations between the reading and quality measures were much stronger for the girls than for the boys, confirming findings reported by Fishco (1966), but the differences between the sexes on the other three writing measures were neither so pronounced nor so consistent. When the experimental and control groups were examined separately, considerably higher correlations were found for the control groups than the experimental groups at all three test periods.

However, when the average reading and writing scores obtained by the experimental and control boys and girls on each grade level were compared, no consistent relationship between the reading and writing scores was evident.

The experimental and control boys and girls at each grade level were ranked from one to four on the reading and each of the four writing measures, offering a total of thirty-two comparisons. On the pretest, the test which was not influenced by the treatment, a group's rank on the reading test could be used to predict its rank on the writing measure on only nine

(i.e., twenty-eight percent) of the thirty-two comparisons. On the other hand, on four (i.e. thirteen percent) of the comparisons, the group recording the lowest average score on the reading measure recorded the highest average score on the writing measure. Nor was a group's growth on the reading measure a good predictor of its growth on the writing measure. The group showing the greatest growth on the reading measure also showed the greatest growth in the writing measure on only sixteen of the thirty-six comparisons.

B. Conclusions

Evidence gathered in the current study does not support the hypothesis that measurable changes in reading ability produce measurable changes in writing ability. Under the conditions described in the current experiment, statistically significant changes on the reading measure did not produce statistically significant changes on any of the four major writing measures at either the grade-nine or the grade-ten level. Only three of the thirty-six null hypotheses tested in the study were statistically significant at the .05 level of confidence. All three of these favored the control group. The small number of differences which achieved significance at the .05 level of confidence must be considered spurious in the light of the large number of hypotheses tested. At the .05 level of confidence, chance alone could be expected to yield "statistically significant" results on two (i.e. five percent) of the thirty-six hypotheses tested.

Furthermore, the changes observed were not, for the

most part, in the hypothesized direction. Although the reading changes generally favored the experimental groups, the writing changes were inconsistent, favoring the experimental group on some comparisons and the control group on others. Only two measures showed consistent changes in the hypothesized direction at one or more testing periods. The fluency measure on the pre-post comparison and the error-reduction measure on the mid-post and pre-post comparisons favored the experimental groups on both the grade-nine and grade-ten levels, but eleven of these twelve comparisons were statistically non-significant.

While correlation coefficients suggested that reading and writing were related skills, other methods of comparing the scores on the reading and writing measures did not support this conclusion. Neither ranking the experimental and control groups on the reading and writing measures nor comparing growth on the two measures suggested a strong relationship between the two skills.

Nor did the results appear to be significantly influenced by the grade level, ability level, or sex of the students. Although random interactions were found between the treatment and the sex of the subject, no significant interactions were found between the treatment and the grade level or the ability level of the subjects.

The experimental design, of course, precludes a definitive statement that measurable change in reading skill does not produce measurable change in writing skill. Failure to reject the null hypotheses can lead only to the conclusion that under the conditions outlined in the current experiment,

measurable change in reading skill <u>did not</u> lead to measurable change in writing skill.

C. Implications

The current study calls into question previous theories about the cause-and-effect relationship between reading skill and writing skill, but this, of course, does not suggest that reading and writing should be considered independent skills in the curriculum.

1. Previous Theories. At least as far back as 1931, when Eurich conducted his pioneering investigation into the relationship between reading skill and writing skill, theorists suggested a cause-and-effect relationship between reading and writing. In the current study the total lack of evidence to support this theory strongly indicates, but, of course, does not prove, that there is no simple cause-and-effect relationship between the two skills. This suggests that a far broader view of the interrelationships among the language arts is required than was envisioned in either the current study or by such previous researchers as Eurich (1931), Mathews, Larsen, and Butler (1945), Schneider (1971), or Glazer (1973). Possible investigations to examine these broader interrelationships are noted in the "Suggestions for Further Research" section which follows.

2. Curriculum. Nothing in the results of the current study can be taken to prove that reading and writing should be taught independently of each other. Although under the

conditions of the current study no causative relationship emerged, correlational data both from the current study and from previous research suggested that reading and writing are, in fact, related skills. On the other hand, the current study was unable to verify the hypothesis that improving reading skill was a potentially effective method of enhancing the quality of student writing.

D. Questions Raised by the Findings

The most important question raised by the current findings involve growth in reading: was there sufficient growth in reading to produce growth in writing? The findings also suggest the need for an assessment of certain features of the experimental design: the environment for the collection of writing samples, the composition topics assigned, the writing measures used in the study, the length of the study, the grade level of the subjects, and the amount of writing instruction and practice the students received. The informal data collected prior to the study, data which appeared to offer promise of improved writing skill as a result of the treatment, must also be re-examined in the light of the findings.

1. Reading Changes. Since the writing changes hypothesized in the current study were based on reading changes, the first questions requiring examination concern the magnitude of the reading changes. Over the six months of the study the

experimental students showed gains in reading superior to those shown by the control students, gains which were statistically significant at the .0013 level of confidence. These gains represented roughly one-half year's growth on the norms accompanying the reading test and did not substantially favor either the grade-nine or the grade-ten samples. Since previous research provides no evidence on the amount of reading change required to produce writing change, it might be suggested that one-half year's growth in reading is simply not adequate to produce measurable writing change. However, other evidence gathered in the study does not support this suggestion. As noted above, neither the direction of the reading and writing changes observed in the study nor comparison of the average reading and writing scores on the initial test indicates a close relationship between the reading and writing scores. This suggests that reading and writing do not have the cause-and-effect relationship hypothesized in this study.

Nonetheless, it is possible that one-half year's growth on the reading measure is simply not adequate to produce writing changes. At this time it is not known how much change in reading is required to generate measurable change in writing, assuming that a cause-and-effect relationship does indeed exist.

2. Obtaining the Writing Sample. In the current study both the environment in which the compositions were written and the topics assigned may have had an influence on the results, but this influence was not expected to be a major factor in

the findings.

a. <u>the environment</u>. While the method of obtaining the writing samples in the current experiment followed methods used by previous researchers, the environment seems to have offered little incentive for the student to produce his best work. In the current experiment, as in previous writing experiments, pretests were not evaluated and returned to the students lest this introduce systematic bias into the experiment. To insure consistency of administration, the investigator rather than the classroom teacher administered all tests. Unfortunately, these features appear to have removed an incentive for the student to produce his best writing and may have been responsible for what appeared, subjectively to the investigator and others, to be the generally pedestrian and uninspired quality of the writing sample as a whole, in contrast with other large samples of student writing from recent years where informal comparisons were possible (e.g., those in the studies of Martin and of Giffen, from 1968 and 1972).

The natural classroom setting provides writing incentives both of a personal nature (the student is communicating with the teacher, a person with whom he has developed rapport) and in the form of extrinsic rewards (the composition is graded and returned to him), neither of which was present in the current experiment. In this study the investigator was in the classroom for only six periods during the school year, three to administer reading tests and three to

administer writing tests, each time in a negative role (bringing work for the students to do), and had no opportunity to develop rapport with the students. Since the students were not graded for their work, it must have become apparent during the second and third testing periods that there was nothing to be gained by writing a good composition, a hypothesis which is supported by the grade-nine student who merely wrote his name on the third paper and left the sheet blank.

The general emphasis on writing in the English teaching profession also appears to have shifted in recent years, possibly as one result of the various back-to-basics movements. An example of this shift can be seen in the thematic topics for _English Journal_. During the 1973-74 school year, when _English Journal_ began organizing issues on themes, only one issue, the December 1973 issue, included the theme "writing," and even in this issue "Focus on Writing" shared the issue with "Best Young Adult Books." During the 1977-78 school year, however, three of the nine issues emphasized writing: "College Students Write" (November 1977), "Teaching Writing" (December 1977), and "Scientific Writing" (April 1978). The earlier lack of emphasis on writing in the profession was probably reflected in classroom practice and might have resulted in a lack of value placed on writing by the students. This, in turn, might have been at least partially responsible for the generally low quality of the writing sample.

Furthermore, the reading and writing portions of the experiment appear to offer different degrees of incentive to the students. On the high school level, there is at the very

least a negative incentive to read: reading does not require as much effort as writing, and is sometimes a welcome relief from other classroom activity. Writing, on the other hand, is a challenging task that places substantial demands on the student. First, writing is difficult work: the student is required to think and to organize, activities which may hold less appeal than simply daydreaming. Second, writing is personal: the student risks rejection by the reader, or at least feels he might, if he makes statements which do not please the reader.

Therefore, in retrospect, it appears that the quality of the writing sample might have been improved by having the classroom teachers administer the tests (with, perhaps, the researcher present as an observer), and by having the writing samples evaluated and returned to the students. The assignment of the topics in six orders of presentation in the current experiment,[1] of course, precluded evaluating and returning the pre- and midtests to the students.

b. <u>the topics</u>. Although the topics used in the current experiment were adapted from the three general areas most frequently chosen by students in Bell's (1971) survey of local student writing preferences, the topics used in the current study appear to be both restrictive and uninspiring. Furthermore, the method of assigning the topics was not one

---

[1]A similar assignment of writing topics was made by both Mellon (1969) and O'Hare (1973) although in these experiments there were only two sets of topics (pre- and posttests) rather than the three in the current experiment.

that would be typical in the natural classroom setting.

Given a free choice of topics, the students in Bell's survey chose the general topics "home," "school," and "society" most often as topics on which to write compositions. However, in Bell's survey it seems reasonable to assume that each writer took a different approach to the topic. In the current study, on the other hand, each student was assigned the same general approach. Furthermore, as evidenced by the number of other topics chosen by Bell's subjects, these topics did not appeal to all students. The current experiment, however, demanded that all students write on the same three topics.

An equally important problem in the current experiment was the lack of preparation the students were given for the writing assignment. The students were not introduced to the topics beforehand, nor were they offered any pre-writing activities, as either of these were thought likely to introduce bias into the experiment. The students were undoubtedly unaccustomed to being given topics in this manner. The usual school writing assignment probably grows out of a class discussion: the students read a poem, a story, or an essay, discuss this in class, and then prepare a written response. Consequently, the students have ideas generated in class to act as a starting point. In the current experiment they had no such help with the assignment, a feature which appears common to writing research. Although class discussions or other stimuli such as films, stories, or poems risk introducing systematic bias into the writing samples by

providing an advantage to one group or another through the kinds of ideas that may be generated in the various classrooms, this might be a less serious problem than the lack of direction offered in the current experiment.

c. implications. While it might be argued that the environment and topics were similar for the experimental and the control groups, these conditions might well have led to an inaccurate assessment of the true writing ability of the students. An accurate assessment of one's ability depends on one's doing one's best. The Olympic runner and the Saturday jogger might well run side-by-side in a race that neither cared about winning.

Although the environment was a problem in the current experiment, it cannot be charged with the total failure to reject the null hypotheses. The pretests, for example, appeared to provide more motivation than either the mid- or posttests did (the experiment offered a novel situation; the students were not aware that they would not receive the compositions back), yet the quality of these compositions, either those of the experimental or the control students, was not appreciably different from that of the other two tests. Nor was there a stronger relationship between the reading and writing measures on the pretest than there was on the posttest.

3. Measures of Reading and Writing Skill. Minor problems arose with the reading and two of the writing measures used in the study, but these problems are unlikely to have had a

substantial influence on the findings. Nevertheless, perhaps many of the apparent anomalies in the current experiment can be attributed to the insensitivity of the measures.

a. the reading measure. As noted in Chapter III, the Davis Reading Test, Series 2, was reviewed very favorably in Buros and was said to be valid from the beginning of grade eight to the end of grade eleven. In the current experiment, however, this span of validity is questionable. On the pretest, while there were no perfect scores, twelve-percent of the grade-ten students scored eighty-eight-percent and higher, the highest, an experimental student, scoring ninety-five-percent. On the following tests, then, these students had very little opportunity to show improvement. This is thought to favor the control students, the students who on the whole showed the fewest gains in reading scores, and suggests a wider disparity between the experimental and control groups than the results reported in this study show.

b. the writing measures. Data gathered in the current study suggest that Diederich's method of rating compositions is effective in achieving high rater reliabilities; however, for experimental purposes, the evaluations obtained using this method may not be sufficiently discriminating. The results of the study also suggest that the Syntactic Density Score is not a valuable tool for evaluating student compositions.

i. Diederich's Method. As reported in Chapter IV, Diederich's (1974) suggested method for rating compositions

not only achieves high rater reliabilities but also tends to compensate for daily fluctuations in ratings. The three-rater method achieved rater reliabilities of .858 for the total sample, but the two primary raters had reliabilities of only .506. This method also minimized daily fluctuations. The lowest reliabilities reported for the two primary raters, .398 on day three, were raised to an acceptable .847 by the use of the third rater. However, as is pointed out in Appendix C, Diederich's method is probably better suited to normal classroom evaluation where the student is being compared with his peers than it is to the experimental situation where the student is being compared with himself. In the classroom situation (or the external examination situation) the objective is to rank a student in relation to his peers. A pretest and a posttest are generally not being compared. In the experimental situation, on the other hand, the student's growth between the pre- and posttests is being measured.

In the current experiment the 54 compositions of one experimental and one control class were randomly intermingled and evaluated during each of the four two-hour evaluation sessions. Following the rating sessions, the three compositions each student wrote were placed in pre-mid-post order and the compositions were re-examined. When the compositions were ordered in this manner rather than randomly intermingled, the grades assigned in the quality evaluation did not appear to indicate careful distinctions in the quality of the papers. Consequently, as discussed in

Appendix C, a twenty-five-percent sample of the compositions used in the quality analysis was re-rated. The papers each student wrote were placed in pre-mid-post order, the treatment identity of the compositions was removed, and raters were asked to estimate the growth evidenced in each set of three compositions. This method also failed to reveal statistically significant differences between the experimental and control groups, and, consequently, was not pursued using a larger sample. However, future researchers might find that this method yields more discriminating evaluations than traditional methods of evaluating compositions do.

ii. The Syntactic Density Score. The published version of the Syntactic Density Score has two serious problems. The first problem, the mathematical scoring anomaly discussed in Appendix D, was discovered and corrected prior to the statistical analysis and, therefore, was not a factor in the results of the current study. The second problem with the measure, however, its heavy reliance on the T-unit, rendered the measure redundant in the current study. The very high correlations ranging from .90 to .94 between the Syntactic Density Score and the T-unit measure suggest that these two measures evaluate almost exactly the same features.[1]

---

[1]These findings confirm those reported by O'Donnell (1976), findings published after the current study was in progress. However, O'Donnell did not use the modified version of the measure used in the current study, and, consequently, differences between the two reports might have been expected.

Considering the extra time and effort required to calculate the Syntactic Density Score when compared with the relatively simple calculation of the T-unit scores, and considering the very minor differences between the two measures, inclusion of the Syntactic Density Score in analysis of writing samples does not appear to be justified.

4. Instruction and Practice. In the current experiment, other than the reading treatment administered in the first two weeks of the experiment, the experimental design did not call for any instruction or practice in reading or writing skills beyond that offered in the normal school program. The experiment hypothesized that improved writing skill would be a natural consequence of improved reading skill,[1] and, consequently, the results raise the question of whether improved reading skill combined with reading practice and/or writing instruction and practice would lead to greater gains in writing skill than either reading practice or writing instruction and practice alone would.

a. reading practice. One traditional theory on the relationship between reading ability and writing skill, as noted in Chapter II, is that wide reading leads to improved writing skill. This theory received empirical support in

---

[1]This was also the hypothesis of previous reading and writing experiments conducted by Eurich (1931), Mathews, et al., (1945), and Schneider (1971). This is apparent from the fact that all three experiments used such exercises as vocabulary exercises and the SRA kits for instruction rather than wide, free reading.

experiments by such researchers as Heys (1962) and Fader and McNeil (1968) who demonstrated that periods of time set aside for free reading resulted in improved writing skill.

Although it appears reasonable to suggest that during the current experiment the subjects received considerably more practice in reading than they did in writing (they would be required to do more reading than writing in their social studies and science courses, for example), perhaps this amount of practice was insufficient to produce changes in writing ability, or, in fact, to produce the optimal changes in reading skill.

It might be hypothesized that improved reading skill provides the foundation for developing writing skill, but that wide reading is required to translate the reading skill into writing improvement. To test this hypothesis, a future investigation might provide free-reading time[1] to both experimental and control subjects but the reading treatment to only the experimental subjects. If growth in reading skill provides the foundation for developing writing skill, the experimental students should show more growth in writing skill

---

[1]It is not known how much free-reading time is required to influence writing skill. Heys (1962), however, devoted approximately 20 percent of the class time over a period of one year to free reading. Elley, et al. (1976), on the other hand, found no significant differences in the compositions of groups which had devoted 40 percent of their English classes to free reading and another 40 percent to literature study when these groups were compared with groups studying either traditional or transformational grammar. As was noted in Chapter II, slight differences favored the groups studying literature at the end of the first year, but slight differences favored the transformational-grammar group at the end of the third year.

than the control students show.

b. writing instruction and practice. Evidence was not gathered in the current study on the amount of writing the subjects did, both because of the logistical problem of collecting the data from the many teachers who would be in contact with the subjects and because writing practice was not a part of the experimental design. Nor is the provincial curriculum an aid in establishing the amount of writing that a typical grade-nine or grade-ten student is required to do. Although the Secondary Language Arts Handbook (1972) offers many suggested writing activities, it suggests no minimum number of compositions that a grade-nine or -ten student should write during the year, nor does it prescribe a specific program in writing instruction beyond "Single and multiparagraph reports, compositions, short stories, speeches, etc. . . . Punctuation" and "Grammar".

The current experiment assumed that improvement in reading skill was the result of an increased ability to process the code that is common to both reading and writing.[1] However, since the encoding process clearly requires skills not required by the decoding process (the ability to generate sentences, the ability to organize thoughts, familiarity with the conventions of the written dialect such as orthography and punctuation, for example) perhaps instruction in the skills

---

[1]The reading treatment used in the study, as reported in Appendix A, was code-oriented.

not common to reading and writing is required.

It might be hypothesized that improved reading skill provides a foundation on which writing instruction may be built. To test this hypothesis a future investigation might provide writing instruction to both the experimental and control subjects but the reading treatment to only the experimental subjects. If growth in reading skill provides the foundation for writing instruction, the experimental students should show more growth in writing skill than the control students show.

5. Length of the Study. Experimental research offers no firm evidence on the length of time required to effect writing change. Burton (1973) suggested that six months appeared to be the minimum time required to bring about statistically significant writing change, noting that most studies of shorter duration failed to find statistically significant differences between the treatment and the non-treatment groups. Although such studies as Gajadharsingh's (1970) are exceptions to this rule, Gajadharsingh's treatment, generative rhetoric, taught sentence-construction skills directly. It was theorized, on the other hand, that the treatment in the current study would elicit an induced skill that would transfer automatically to the skill of writing. Perhaps this transfer requires more than the six months that the study allowed. Studies in the social sciences reported by such researchers as Sullivan, et al. (1975) suggests that teaching values as opposed to teaching skills requires considerably longer than one year.

On the other hand, charging the failure to reject the null hypotheses in the current study to insufficient time would be more plausible if the changes in writing skill were in the hypothesized direction. Improvements in writing skill in the current study, however, were randomly divided between the experimental and control groups.

Nevertheless, perhaps a longitudinal study of two years or more, a study almost precluded by scheduling difficulties in the local public high schools, would indicate writing changes that were delayed beyond the six months of the current study.

6. Grade Level of the Experiment. Experimental research offers no firm evidence on the optimal grade level for experiments in writing. Potter (1966), choosing grade-ten students to examine structural features in writing, suggested that somewhere between grade eight and grade eleven students' writing ceases to be "speech writ down" and begins to assume some of the characteristics of the prestige written dialect. Others have chosen various grade levels for similar reasons: Mellon (1969) chose grade seven because "it represents a time just in advance of the high school years, when the most noteworthy growth of syntactic fluency normally occurs among the students"; Bateman and Zidonis (1966) chose grade nine because it was at this level that grammar study was customarily the most concentrated; O'Hare (1973) followed Mellon's choice of grade levels but felt that less complex sentence-combining

[illegible]

exercises could be used as early as the primary level.

If the grade-nine and -ten levels represent a state of change, perhaps these levels are not ideal levels on which to conduct writing experiments. In the current experiment the investigator felt that the grade level where the students' writing was already in a state of change might be an opportune time to influence the direction of this change. Perhaps an experiment on the twelfth-grade level when student writing patterns are said to be more stable or on the seventh-grade level where the patterns are just beginning to develop would produce different results.

7. Data Collected Prior to the Study. The results of the current study failed to support either previous findings on the relationship between reading and writing or the informal local data collected prior to the study. As mentioned in Chapter II, both considerable correlational data reported by previous researchers and informal local data suggested that students who received the reading treatment would show more improvement in writing than students who had not received the treatment would. The failure to reject the null hypotheses in the current study raises the question of why the informal data showed promise that was not realized in the experiment. One possible answer lies in the characteristics of the pilot data and another in the environmental factors and writing practice which may have influenced the quality of the pilot data.

a. characteristics of the pilot data. The local pilot

data were clinical rather than experimental and had two shortcomings: first, much of the data consisted of professional opinion rather than "objective" evidence; and, second, the compositions analyzed were few in number and lacked statistical controls. However, the five teachers who used the reading treatment with their students and provided the testimonials all voiced opinions consistent with each other. All claimed a post-treatment improvement in the quality of the compositions written by their students.

The second problem with the pilot data is the lack of experimental controls. The pre- and post-treatment book reports written by the high school students were few in number (only three pairs of compositions) and lacked controls; the post-secondary school sample, although somewhat larger (twelve students) also lacked controls. Nonetheless, in both samples the students who had taken the reading treatment showed improvement, both in the quality of the compositions and on the syntactic maturity measures.

b. writing practice. A feature that all the pilot data had in common that was not necessarily shared by the data in the experiment was the amount of writing practice the students received. As noted in Section 4 above, there is no indication as to the amount of writing that the classes involved in the experiment did. On the other hand, the students providing the pilot data wrote frequently. The teacher who taught the secondary students reported that her students wrote a composition each week. The students in the

post-secondary sample wrote even more often than this although under different circumstances. The adult students in this sample, most of whom did little recent writing before enrolling in the reading course, wrote two reactions to articles, stories, or novels each week, one each class period. Perhaps this amount of practice would have led to writing improvement regardless of the treatment, although the classic study by Buxton (1959) would suggest that this is not likely.

c. environment. In the pilot data the students were writing to a teacher with whom they had developed rapport and in both situations students were writing on subjects they had read about. Furthermore, in both situations the compositions were evaluated and returned to the students. As noted in Section 3 above, the compositions written in the current experiment were neither preceded by pre-writing activities nor returned to the students following evaluation.

E. Suggestions for Further Research

Although replication of studies which fail to find significant differences favoring the treatment groups is often recommended on the basis of the rationale on which the study was based and the previous research which suggested the hypothesized results, it is the opinion of the current investigator that simple replication of the current study, even altering such design features as the environment for

collecting the writing sample and the composition topics assigned, would yield results very similar to those reported here. Nonetheless, there are certain major design features that may be altered to produce different results from those reported in the current experiment.

1. Reading Practice. As discussed in Section D4a above, reading practice was not included in the design of the current study. Previous studies by Heys (1962) and Fader and McNeil (1968) have suggested that mere reading practice leads to improved writing skill. A possible avenue of future research would be to determine if reading treatment followed by reading practice is a more effective vehicle for improving writing skill than reading practice alone is.

2. Writing Instruction and Practice. As discussed in Section D4b above, writing instruction and practice were not included in the design of the current experiment. A possible avenue of future research would be to determine if reading treatment followed by writing instruction and practice is a more effective vehicle for improving writing skill than writing instruction and practice alone is.

3. The Length of the Study. As noted in Section D5 above, it is not known if writing changes occurred beyond the six months of the current study. Although the evidence gathered in the current study suggested that changes in reading skill had virtually no effect on growth in writing skill, perhaps these changes would have appeared one or two years later,

[illegible]

[illegible]

[illegible]

[illegible]

after the subjects had been exposed to more reading and writing in their school work.

4. <u>The Grade Level of the Study</u>. As noted in Section D6 above, the optimal grade level on which to conduct writing experiments is not known. A future study might investigate the effect of improved reading skill on the writing of students either below the grade-nine level or above the grade-ten level where the writing patterns of students are said to be more stable.

5. <u>A Study Using Available Resources</u>. Diederich (1974) suggested that studies of a much broader scope could be carried out if researchers used evaluations already being done in the schools. One objection to this suggestion is that the grades assigned in school generally lack the reliability demanded of experimental studies. In Alberta, however, the recent province-wide evaluation of reading and writing skills on the grade-twelve level makes Diederich's suggestion practicable. These tests are administered and analyzed under rigorous conditions. What follows would apply equally, of course, in any setting in which annual testing of writing is being performed.

The ease of administering the reading treatment in the current study combined with the availability of data through the province-wide examinations suggests an economical method of replicating the current study, a replication which would eliminate both the writing

incentive and writing environment problems discussed above. This replication would also allow a longitudinal study of three or more years. Such a study could embody the following features:

a. the sample. Because the reading treatment in the current study may be administered in approximately forty minutes and may be administered in groups ranging up to four in number, subjects could be drawn from a variety of schools. Therefore, the population from which the experimental and control students are randomly chosen could be as large as the entire population of senior-high-school students in the twelve Edmonton public senior high schools.

b. administration of the treatment. Assuming a sample size of 100 experimental and 100 control students, approximately the size of the sample in the current study, one might expect approximately eight experimental students in each of Edmonton's twelve public senior high schools. Therefore, the administrator of the treatment could expect to spend one-half day in each school or approximately six days to complete the administrations.

A longitudinal study of three years could be carried out by administering the treatment to grade-ten, -eleven, and -twelve students during one year. The results of the reading and writing examinations could then be collected for each of these groups when the group completed grade twelve.

c. measurement. With random samples drawn from such a

large population, there would be no need for pretesting[1] and no need to control instructional variables. The environment in which the examinations are written, the incentive the examinations provide for the student to produce his best work, and the methods of scoring the examinations should combine to provide highly reliable and valid measures of the subjects' abilities in reading and writing. In addition, the investigator could secure copies of the compositions to examine structural features such as T-unit length.

F. Epilogue: Re-examination of the Study

This study began with an idea which appeared obvious to the investigator: reading and writing skills are so closely related that anything which affects one skill will almost certainly affect the other. Much as the telegrapher who is skilled at receiving the Morse code should be expected to be skilled at sending the code, the person who is skilled at receiving information from the printed page should also be skilled at producing written work. It would appear, then, that improvement in one skill would lead to improvement in the other. This relationship appeared so obvious that it did not seem open to empirical study. Surely

---

[1] Data already available in the schools such as previous marks in English classes and scores on standard measures such as IQ tests might be used to examine the initial equivalence of groups if such an examination were deemed necessary.

a question as basic as this would have been answered conclusively long ago. However, a review of the literature failed to reveal any empirical evidence that enhancing reading skill would result in a consequent enhancement of writing skill. A number of investigators who had attempted to examine this relationship failed to produce measurable reading changes and consequently were unable to assess writing changes which depended on these reading changes. However, some evidence did suggest that the converse was in fact possible: teaching writing skills did result in reading improvement. The literature also provided indirect evidence that reading and writing skills had a causative relationship. There were, for example, the well-known and well-documented interrelationships among the language arts which suggested that reading and writing were related skills. In addition, the literature provided correlational evidence and evidence showing that the code methods of teaching beginning reading could influence a student's skill in writing, especially Smith (1968).

Local pilot data also suggested that improving reading skill would result in an enhancement of writing skill. Teachers who had used the reading treatment in their classrooms reported that their student's writing had improved.[1]

---

[1]The teachers volunteered the discussion on writing changes without being asked specifically about writing. They were not given a questionnaire with the heading "writing changes" on it, nor were they asked to comment on writing changes. They were simply asked to note any changes they had

These changes in writing skill, then, appeared to be obvious, not the type that would require sophisticated measuring devices to detect. The compositions in the pilot data, though few in number, also suggested that the reading treatment would influence writing skill. Not only were the post-treatment compositions of obviously higher quality than the pre-treatment compositions, but the differences between the two were also quantifiable on both the syntactic density and T-unit measures.

Because the S.O.S. Reading Technique had been shown in experimental settings to improve students' reading skill, the technique offered the opportunity to examine the question of what writing changes would result from improved reading skill. The conclusion seemed obvious: the writing of the treatment groups would improve. Therefore, the major question appeared to be not if the writing of the treatment group would improve, but what kinds of writing improvement would be evident. Would, for example, the major improvements be most obvious on the quality, syntactic maturity, or sentence error measures? Would the changes be the same on one grade level as on another? Would they be the same for the boys and girls? Would the ability level of the student influence the changes? Would the writing changes follow the reading changes promptly or would they be delayed and appear more

---

observed in the work of their students following the administration of the reading treatment (E. Martin, 1976, pp. 6-7).

strongly later in the experiment?

Nonetheless, prior to the experiment it seemed possible, though not likely, that such hazards as the problems of assessing writing skill or the time allotted for the experiment would preclude finding statistically significant differences between the experimental and control groups. If the differences were in the hypothesized direction but not distinct enough to reach statistical significance, replication of the study would be suggested. If such trends were shown, features of the experimental design which appeared to account for the weakness of the results might be altered and the experiment replicated on the basis that certain aspects of the results showed promise.

The results of the current study, however, were completely unexpected and appear to offer no promise of significant results if features of the design are altered. Despite substantial and statistically significant changes on the reading measure, no trends were found on any of the four writing measures to suggest that reading and writing have a causative relationship. Changes on the four major writing measures appeared to operate in a purely random manner, with slight fluctuations favoring the control group on one comparison and the experimental group on another. The probabilities found in the study were generally above .50, suggesting that these differences were indeed slight. The subsidiary analysis of major errors was the only writing analysis to yield trends in the hypothesized direction, and

even these trends were not totally consistent, suggesting only that further study may be warranted. The correlations found in the study, however, did confirm previous research on the relationships between reading and writing skills. Correlations between the reading and writing quality measures ranged as high as .68 for the girls in the study. While correlation is obviously not the same as causation, it is still difficult to believe that an appreciable change in reading would produce no effect on writing. However, since the gains on the reading measure appeared to have absolutely no effect on the writing sample, one is left, it seems, with two possible alternatives: 1. that in some way which is not apparent to the investigator the study is flawed and has not given a true assessment of the students' reading and/or writing skills, or 2. that the hypothesis of a causative relationship between reading and writing skills, a hypothesis which seemed so obvious at the onset of the study, is in some serious way in error.

1. The Experimental Design and the Reading and Writing Measures. Although, as noted above, minor problems were experienced with the topics chosen for the writing samples and the difficulties of June testing, the experimental design appears to be sound. The design used was a modification of the classic pre-post design with random assignment of groups to treatments. The statistical analysis was appropriate for the nature of the data, and the time allotted seemed adequate to expect trends in the hypothesized

direction, if not statistically significant changes.

The reading results obtained in the study appear to be valid. The Davis Reading Test is widely regarded as one of the best tests available for measuring reading skill on the grade-nine and -ten levels. Analysis of the reading data for the six months of the study shows that the changes were both statistically significant and substantial. Statistical results suggest that differences as great as those observed between the experimental and control groups on the reading measure could be attributed to some factor other than the treatment 13 times in 10,000, a probability that is very slight. Inspection of the raw data confirms the statistical analysis. The experimental students gained over three raw-score points more than the control students did over the six months of the experiment. Furthermore, these gains were distributed fairly evenly among the grade nines and tens and the girls and boys, and were most pronounced during the first three months of the experiment. They appear, then, not to be the result of gains shown by one small segment of the experimental group. Nor was there an appreciable delay in the reading changes following the administration of the treatment, a delay which could suggest that the writing results might be expected sometime following the six months of the experiment. In any case, the reading results reported in the current study merely confirmed those reported in two previous studies. Although further experimentation may in fact provide evidence contrary

to that reported here, based on the present data it does not appear that the reading results can simply be dismissed as spurious.

Although the writing results are not supported by such a prestigious measure as the Davis Reading Test, they do not appear open to serious challenge. The major writing measures approached the compositions from two different perspectives: the subjective measure of overall quality and the objective measures of syntactic maturity. These measures appear to incorporate the best current practice in the measurement of writing skill. It might be suggested that because of the notorious problems associated with measuring writing skill, the measures simply failed to assess writing changes. This, however, would fail to take account of the many experiments which have demonstrated measurable writing change, and impressively so, relying on measures which in many cases appear less thorough and less objective than those used in the current study. O'Hare (1973), for example, reported impressive changes on the writing quality measure using rating techniques which appear less thorough and statistical techniques which appear less discriminating than those used in the current study. The quality rating obtained in the current study was done under strictly controlled conditions, was preceded by rater training sessions, and was supported by high interrater reliabilities. Analysis of covariance, used in the current study to determine statistical differences between the experimental and control

groups was more discriminating than the non-parametric statistic used by O'Hare. Furthermore, a sample of the compositions rated in the quality analysis was re-rated in an attempt to obtain even more precise ratings, but this, too, failed to reveal differences between the experimental and control groups. The other writing measures used in the study, especially the T-unit measure, have been shown in previous studies to be good indicators of writing skill and writing maturity. It appears, then, that the writing results, like the reading results, cannot be dismissed as simply spurious.

Both the length of the study and the amount of reading change shown by the experimental students might also be considered possible reasons for the absence of writing changes in the experiment. It is not known, for example, if the experimental students would show improved writing skill two years or five years following the administration of the treatment. This possibility seems remote, however, since although the reading changes were prompt (the effects were statistically significant three months following the administration of the treatment) there were virtually no apparent writing changes six months following the administration of the treatment. And there is, of course, no indication of how many units of "cause" are required to produce one unit of "effect." It is possible that the average three raw-score units of reading change in the current study were not large enough to produce any effects on the writing of the students. Perhaps ten such units are required to produce writing changes. This, again, would be a more reasonable explanation if trends

were found in the hypothesized direction.

2. Re-examination of the Theory Underlying the Study. If one accepts the reading and writing results as valid and concedes that the experimental design is not in some way seriously flawed, and there appears to be little reason not to accept these premises, what are the implications of the study? The complete lack of evidence in the current study to support the view of a causative relationship between reading and writing skills and the lack of any trends in the hypothesized direction suggest (but, of course, do not prove) that a direct causative effect of reading skill on writing skill does not in fact exist. This, of course, raises questions about the integrated approach to teaching reading and writing. If reading and writing are in fact discrete skills, as the current study suggests they are, then perhaps they must be taught separately if maximum benefit is to be derived. This question ought to be investigated carefully by those who advocate the integration of reading and writing in the classroom.

The results of the current study also raise questions about the correlations between reading and writing skills. While the current study suggests that there is no direct causative relationship between reading and writing skills, the correlational data suggest that these skills are not totally independent of each other. The sheer weight of correlational evidence from a variety of sources including this study is too large to be simply dismissed as spurious.

One possible explanation of the correlations between reading and writing skills is that writing is indirectly influenced by reading. Through immersion in reading the student absorbs both the vocabulary and sentence structures of the written dialect and also increases his ability to deal with language and thought. This reading, then, would be reflected in both his reading skill and his writing skill. To date, evidence on the efficacy of large-scale reading practice to improve writing skill is inconclusive: researchers such as Heys (1962) and Fader and McNeil (1968) reported that wide reading had a positive influence on writing skill, but Elley, et al., (1976), in a much more thorough study than the previous two, reported that wide reading did not influence writing skill. It may be, however, that if such programs are eventually demonstrated to hold important curricular promise, then a means of improving reading skill at the onset of the programs might well have the ultimate effect, though quite indirectly, of improving student writing. The widely-reported correlations between the two skills combined with the suggestion in the present study that reading skill *per se* does not influence writing skill appears to add pertinence to such a line of enquiry.

Another possible explanation of the correlational data is that reading and writing are both influenced by some third factor, general language proficiency, for example. Loban (1966) has offered research support for this suggestion. Loban noted that the student who comes from a

high socio-economic background and is generally accustomed to hearing complex sentence structures and rich vocabulary in the home, enters school with an advantage in dealing with the written dialect. With this head start, the student with the rich language background is better able to cope with the printed page than is the student lacking such a background. If both reading and writing skills are influenced by language abilities acquired in the home, it appears reasonable that correlations would be found between these two skills. Although this suggestion appears to doom the student with a weak language background to poor writing skills, it does not, in fact, suggest that writing skills cannot be improved. Results reported by such investigators as O'Hare (1973), Miller and Ney (1968), and Gajadharsingh (1970) using various forms of sentence-combining exercises, Obenchain (1971) using lessons in sentence and paragraph development, and Buxton (1959) using theme revision belie any such suggestion. The common characteristic of these studies, however, is that the researchers taught writing or aspects of writing skill. Contrasted with the paucity of writing changes in the current study, these results suggest that one should not expect to teach writing by indirect methods: if writing is to improve, it should be taught. To paraphrase Christensen (1963): we must not merely expect students to become better writers; we must teach them to write.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abrams, Arnold G., "The Relation of Listening and Reading Comprehension to Skill in Message Structuralization," Journal of Communication, vol. 16 (June, 1966), pp. 116-25.

Ackerman, S.A., "An Analysis of First Grade Children's Writings as they Reflect Three Different Methods of Teaching Initial Reading," unpublished doctor's dissertation, Chico State College, 1969, in Lenore Sandel (ED 019 144).

Adams, Sharon, "Reading, Writing on the Downgrade, Teachers Find," The Edmonton Journal, (February 21, 1976), p. 17.

Akers, Charles W., "Effects of a Supplemental Program in Structural Analysis Upon Reading Scores of Children Deficient in Grammatic Closure," unpublished doctor's dissertation, Ohio University, 1976, in Dissertation Abstracts International, vol. 37 (1976), p. 4598A.

Allen, Robert L., "Better Reading Through the Recognition of Grammatical Relations," The Reading Teacher, vol. 18 (1964), pp. 194-98.

Andreach, James R., "The Use of Models to Improve Organizational Techniques in Writing," unpublished doctor's dissertation, Yeshiva University, 1975, in Dissertation Abstracts International, vol. 36 (1975), p. 4980A.

Applebee, Arthur N., "Writing and Reading," Journal of Reading, vol. 20 (March, 1977), pp. 534-36.

________, Tradition and Reform in the Teaching of English: A History, Urbana, Illinois, National Council of Teachers of English, 1974.

Artley, A. Sterl, "Research Concerning Interrelationships Among the Language Arts," Elementary English, vol. 27 (December, 1950), pp. 527-37.

________, "A Study of Certain Factors Presumed to be Associated with Reading and Speech Difficulties," Journal of Speech and Hearing Disorders, vol. 13 (December, 1948), pp. 351-60.

________, Gertrude H. Hildreth, Agatha Townsend, Althea Beery, and Mildred A. Dawson, Interrelationships Among the Language Arts, Research Bulletin of the National Conference on Research in English, 1954.

Ashida, Margaret E., and Leslie T. Whipp, "A Slide Rule Composition Course," College English, vol. 25 (October, 1963), pp. 18-22.

Auguste, Joanne, and F.B. Nalven, "How Lasting are the Effects of ITA vs. TO Training in the Development of Children's Writing?" Research in the Teaching of English, vol. 6 (Spring, 1972), pp. 17-20.

________, "ITA and TO Training in the Development of Children's Creative Writing," Research in the Teaching of English, vol. 3 (Spring, 1969), pp. 178-80.

Baker, William D., "An Investigation of Characteristics of Poor Writers," College Composition and Communication, vol. 5 (1954), pp. 23-27.

Barbig, Evelyn, "An Exploration of Growth in Written Composition to Determine the Relationship of Selected Variables to Poor Writing in Grades 9 and 12," unpublished doctor's dissertation, University of Tennessee, 1968, in Dissertation Abstracts International, vol. 29 (1968), p. 2140A.

Bateman, Donald, and Frank Zidonis, The Effect of a Study of Transformational Grammar on the Writing of Ninth and Tenth Graders, National Council of Teachers of English Research Report No. 6, Champaign, Illinois, National Council of Teachers of English, 1966.

Becker, Samuel L., "Relationships Among the Communication Skills, Central States Speech Journal, vol. 14 (November, 1963), pp. 258-63.

________, Carl A. Dallinger, Harry H. Crosby, and David Gold, Communication Skills: An Experiment in Instructional Methods, Iowa City, University of Iowa Press, 1958.

Belanger, Joseph F., "Calculating the Syntactic Density Score: A Mathematical Problem," Research in the Teaching of English, vol. 12 (May, 1978), pp. 149-53.

Bell, James B., "A Study of the Written Composition Interests of Senior High School Students," unpublished doctor's dissertation, University of Oregon, 1971.

Biggs, Bernice P., "Construction, Validation, and Evaluation of a Diagnostic Test of Listening Effectiveness," Speech Monographs, vol. 23 (March, 1956), pp. 9-13.

Binzer, Patricia J., "The Effect of Direct Instruction in Comprehension Through Listening on Reading Comprehension," unpublished doctor's dissertation, State University of New York at Albany, 1974, in Dissertation Abstracts

International, vol. 34 (1974), p. 938A.

Bippus, Anne Clark, "The Relationship of the Quality of Students' Written Language, Productivity of Writing, and Reading Comprehension in Grades Four and Six," unpublished doctor's dissertation, University of Virginia, 1977, in Dissertation Abstracts International, vol. 38 (1977), p. 3993A.

Blau, Harold, "Written Composition and Oral Discourse," English Journal, vol. 57 (March, 1968), pp. 369-71.

Blewett, Thomas, "An Experiment in the Measurement of Listening at the College Level," unpublished doctor's dissertation, University of Missouri, 1949, in Dissertation Abstracts International, vol. 9 (1949), p. 87.

Blount, Nathan S., "Research in Teaching Literature, Language, and Composition," in Robert M.W. Travers (ed.), Second Handbook of Research on Teaching, Chicago, Rand-McNally, 1973, pp. 1084-97.

Blumenfeld, J.P., and G.R. Miller, "Improving Reading Through Teaching Grammatical Constraints," Elementary English, vol. 43 (1966), pp. 752-55.

Bond, G.L., The Auditory and Speech Characteristics of Poor Readers, Teachers College Contributions to Education, No. 657, New York, Teachers College, Columbia University, 1935.

Braddock, Richard, "English Composition," in Robert L. Ebel (ed.), Encyclopedia of Educational Research, 4th edition, Toronto, Collier-Macmillan Canada Limited, 1969, pp. 443-61.

________, Richard Lloyd-Jones, and Lowell Schoer, Research in Written Composition, Champaign, Illinois, National Council of Teachers of English, 1963.

Broudy, Harry S., and John R. Palmer, Exemplars of Teaching Method, Chicago, Rand-McNally, 1965.

Brown, Clark, "The Devil's Syllabus," College Composition and Communication, vol. 26 (December, 1975), pp. 365-67.

Brown, James I., "A Comparison of Listening and Reading Ability," College English, vol. 10 (1948), pp. 105-07.

Buckingham, B.R., "Language and Reading - A Unified Program," Elementary English Review, vol. 17 (March, 1940), pp. 111-16.

Buros, Oscar K. (ed.), The Sixth Mental Measurements Yearbook, Highland Park, New Jersey, The Gryphon Press, 1965.

Burrows, Alvina T., Margaret B. Parke, Neal R. Edmund, John J. Deboer, Thomas D. Horn, Virgil E. Herrick, and Ruth G. Strickland, Children's Writing: Research in Composition and Related Skills, Champaign, Illinois, National Council of Teachers of English, 1961.

Burton, Dwight L., "The Troubled Dream," Research in the Teaching of English, vol. 7 (Fall, 1973), pp. 160-90.

Bushnell, Paul, An Analytical Contrast of Oral with Written English, Teachers College Contributions to Education, No. 451, New York, Teachers College, Columbia University, 1930.

Buxton, E.W., "An Experiment to Test the Effects of Writing Frequency and Guided Practice upon Students' Skill in Written Expression," Alberta Journal of Educational Research, vol. 5 (June, 1959), pp. 91-99.

________, "Teaching Writing in Canada," in James R. Squire, A Common Purpose, The Teaching of English in Great Britain, Canada, and the United States, Champaign, Illinois, National Council of Teachers of English, 1966.

Calhoun, J.L., "The Effect of Analysis of Essays in College Composition Classes on Reading and Writing Skills," unpublished doctor's dissertation, Boston University, 1971, in Dissertation Abstracts International, vol. 32 (1971), p. 1971A.

Calvert, K.H., "K-Ratio (Kernal Structure) Index, An Investigation of the Relationship between Syntactic Maturity of Oral Language and Reading Comprehension," 1971, (ED 091 722).

Campbell, Daisy B., "A Study of Two Methods of Teaching English in a Community College Setting," unpublished doctor's dissertation, Virginia Polytechnical Institute and State University, 1976, in Dissertation Abstracts International, vol. 37 (1976), p. 4808A.

Campbell, James R., "The Effect of Oral Language on Reading Achievement," unpublished doctor's dissertation, Lehigh University, 1973, in Dissertation Abstracts International, vol. 34 (1973), p. 5575A.

Campbell, Mary Lyn, "An Investigation of the Relationship between Secondary Generative and Receptive Communication Skills at the College Freshman Level," unpublished

doctor's dissertation, University of Southern Mississippi, 1976, in Dissertation Abstracts International, vol. 37 (1976), p. 5655A.

Carner, Rebecca L., "The Relative Effectiveness of the ITA and TO on Reading, Spelling, and Writing Achievement of First and Second Grade Children," unpublished doctor's dissertation, University of Miami, 1971.

Chall, Jeanne S., Learning to Read: The Great Debate, New York, McGraw-Hill, 1967.

Christensen, Francis, "A Generative Rhetoric of the Sentence," College Composition and Communication, vol. 14 (October, 1963), pp. 155-62.

Coffman, William E., "Reading Tests and Reviews," in O. Buros (ed.), Sixth Mental Measurements Yearbook, Highland Park, New Jersey, The Gryphon Press, 1965, pp. 291-92.

Combs, Warren E., "Sentence-Combining Practice Aids Reading Comprehension," Journal of Reading, vol. 21 (October, 1977), pp. 18-24.

Cook, Desmond L., "The Impact of the Hawthorne Effect in Experimental Designs in Educational Research," U.S. Office of Education Co-operative Research Project, No. 1757 Columbus, Ohio State University, 1967, (ED 021 308).

________, and Joan King, "A Study of the Hawthorne Effect in Educational Research," Research in the Teaching of English, vol. 2 (Fall, 1968), pp. 93-98.

Cooper, C.R., "Students Learn to Read Fiction by Writing It," California English Journal, vol. 9 (April, 1973), pp. 7-17.

________, and Lee Odell, "Considerations of Sound in the Composing Process of Published Writers," Research in the Teaching of English, vol. 10 (Fall, 1976), pp. 103-15.

Corbett, Edward P.J., "The Theory and Practice of Imitation in Classical Rhetoric," College Composition and Communication, vol. 22 (October, 1971), pp. 243-50.

Crews, Ruthellen, "A Linguistic Versus a Traditional Grammar Program of Effects on Written Sentence Structure and Comprehension," Educational Leadership, vol. 5 (1971), pp. 145-49.

D'Angelo, Frank J., "Imitation and Style," College Composition and Communication," vol. 24 (October, 1973), pp. 283-90.

D'Angelo, John L., "Predicting Reading Achievement in a Senior High School from Intelligence, Listening, and Informative Writing," unpublished doctor's dissertation, University of Pittsburgh, 1977, in Dissertation Abstracts International, vol. 38 (1977), p. 2027A.

Davis, Frederick B., and Charlotte Croon Davis, Davis Reading Test: Manual, Series 1 and 2, New York, The Psychological Corporation, 1962.

Davis, Irene P., "The Speech Aspects of Reading Readiness," Seventeenth Yearbook of Department of Elementary School Principals, vol. 17 (July, 1938), pp. 282-89.

Davis, James E., "The Language Laboratory as a Medium and Model in a Talk-Write Composition Course," NALLD Journal, vol. 9 (Winter, 1975), pp. 16-19.

Dawson, Mildred A., "Interrelationships between Speech and Other Language Arts Areas," in A. Sterl Artley, et al., Interrelationships Among the Language Arts, Champaign, Illinois, National Council of Teachers of English, 1954, pp. 24-35.

Denson, Teresa A., "An Experimental Study of the Effect of a Listening Skills Training Program in Kindergarten on the Development of Pre-Reading Skills," unpublished doctor's dissertation, University of Southern California, 1973, in Dissertation Abstracts International, vol. 34 (1973), p. 3986A.

Devine, Thomas, "Listening and Reading," paper presented to the annual meeting of the Reading Association of Ireland, September, 1976, (ED 132 523).

DeVries, Ted, "Reading, Writing Frequency, and Expository Writing," Reading Improvement, vol. 7 (1970), pp. 14-15.

Diederich, Paul B., Measuring Growth in English, Champaign, Illinois, National Council of Teachers of English, 1974.

________, The Problem of Grading Essays, Princeton, New Jersey, ETS 1957, (mimeographed).

________, Joe French, and Sydell Carlton, "Factors in Judgments of Writing Ability," Research Bulletin, (Educational Testing Service RB 61-15), 1961.

Donelson, K.L., "Variables Distinguishing Between Effective and Ineffective Writers at the Tenth Grade," Journal of Experimental Education, vol. 35 (Summer, 1967), pp. 37-41.

Dow, Clyde W., and Stephen Papp, "The Relation of Reading Ability and Language Ability to Spelling Ability," Speech Monographs, vol. 10 (1943), pp. 107-08.

Downing, John, "The Effects of the Initial Teaching Alphabet on Young Children's Written Composition," Educational Research, vol. 9 (February, 1967), pp. 137-44.

Drumm, Irma D., "The Effect of Instruction in Oral Language-Experience Activities, Synonym Generation, and Kernal Sentence Expansion on Language Development and Reading Achievement," unpublished doctor's dissertation, Temple University, 1976, in Dissertation Abstracts International, vol. 37 (1976), p. 2033A.

Dykes, Margie G., "An Investigation of the Relationship Between Two Models of Instruction in Grammar and Improvement in Reading Comprehension," unpublished doctor's dissertation, University of Kansas, 1976, in Dissertation Abstracts International, vol. 37 (1976), p. 5020A.

Ebel, Robert L. (ed.), Encyclopedia of Educational Research, 4th edition, Toronto, Macmillan, 1969.

Elkins, D., "An Experimental Investigation of Intensive Reading and Intensive Writing to Improve Composition at the Eighth Grade Level," an unpublished doctor's dissertation, American University, 1968, in Dissertation Abstracts International, vol. 29 (1968), p. 1162A.

Elley, W.B., I.H. Barham, H. Lamb, and M. Wyllie, "The Role of Grammar in a Secondary School English Curriculum," Research in the Teaching of English, vol. 10 (Spring, 1976), pp. 5-21.

Eurich, Alvin C., "The Reading Abilities of College Students: An Experimental Study," College Problems Series, Minneapolis, Minnesota, University of Minnesota Press, 1931.

Evanechko, Peter, Lloyd Ollila, and Robert Armstrong, "An Investigation of the Relationships Between Children's Performance in Written Language and Their Reading Ability," Research in the Teaching of English, vol. 8 (Winter, 1974), pp. 315-26.

Evans, Ronald U., "The Relationship of Reading Comprehension to Written Production of Transformationally Lengthened Prose," Journal of Reading Behavior, vol. 9 (Fall, 1977), pp. 295-99.

Fader, Daniel N., and Elton P. McNeil, Hooked on Books: Program and Proof, New York, G.P. Putnam's Sons, 1968.

Fagan, William, "Transformations and Comprehension," The Reading Teacher, vol. 25 (1971), pp. 169-72.

Fenstermaker, John J., "Literature in the Composition Class," College Composition and Communication, vol. 28 (February, 1977), pp. 34-37.

Ferguson, George A., Statistical Analysis in Psychology and Education, Toronto, McGraw-Hill, 1971.

Fishco, D.T., "A Study of the Relationship Between Creativity in Writing and Comprehension in Reading of Selected Seventh Grade Students," unpublished doctor's dissertation, Lehigh University, 1966.

Fisher, K.D., "An Investigation to Determine if Selected Exercises in Sentence-Combining Can Improve Reading and Writing," unpublished doctor's dissertation, Indiana University, 1973, in Dissertation Abstracts International, vol. 34 (1973), p. 4556A.

Folta, Bernarr, "A Comparative Study of the Syntax in Speech and Writing of Grade One Students Using the ITA and Students Using TO, A Final Report," February, 1968, (ED 017 510).

Fosveldt, D.R., "Development of a Scale for Evaluation of High School Student Themes," Montana Education, vol. 39 (November, 1962), pp. 20-21.

Frye, Northrup (ed.), Design for Learning, Toronto, University of Toronto Press, 1962.

Fuller, Katherine M., "An Investigation of the Relationship Between Reading Achievement and Oral and Written Language of Students Enrolled in Reading and English Classes at Gainsville Junior College," unpublished doctor's dissertation, University of Georgia, 1974, in Dissertation Abstracts International, vol. 35 (1974), p. 6692A.

Furness, Edna L., "Parallels in Speaking and Writing," Education, vol. 80 (1960), pp. 265-66).

Fyfe, T.W., "A Study of the Effects of the Initial Teaching Alphabet on Written Composition at Primary III Level," unpublished thesis, St. Andrews University, 1965, (ED 032 295).

Gajadharsingh, Joel L., "A Study of the Effects of Instruction in the Rhetoric of the Sentence on the Written Compositions of Junior High School Students," unpublished doctor's dissertation, University of Alberta, 1970.

Gatlin, Eugene S., Jr., "The Extent of the Relationship Between the Maturity of Oral and Written Extemporaneous Compositions in the Language of Community College Freshmen," unpublished doctor's dissertation, University of South Carolina, 1974, in Dissertation Abstracts International, vol. 35 (1974), p. 6534A.

Gay, Carol, "Reading Aloud and Learning to Write," Alberta English, vol. 16 (Fall, 1977), pp. 4-10; reprinted from The Elementary School Journal, (November, 1976).

Gibbons, Helen, "Reading and Sentence Elements," Elementary English Review, vol. 18 (February, 1941), pp. 42-46.

Gibson, J.W., C.P. Gruner, R.J. Kibler, and F.J. Kelly, "Qualitative Examination of Differences and Similarities in Written and Spoken Messages," Speech Monographs, vol. 33 (1966), pp. 444-45.

Giffen, Donald P., "The Relative Difficulty of Selected Syntactic Structures of English as Revealed by a Cloze Test," unpublished doctor's dissertation, University of Alberta, 1972.

Glazer, Joan I., "The Effect of Literature Study on the Ability of 4th and 6th Grade Pupils to Create Written Stories," National Institute of Education, 1973, (ED 095 538).

Gleason, H.A., Jr., Linguistics and English Grammar, Toronto, Holt, Rinehart, and Winston, 1965.

Goldstein, Harry, Reading and Listening Comprehension at Various Controlled Rates, Teachers College Contributions to Education, No. 281, New York, Teachers College, Columbia University, 1940.

Golub, Lester S., "Syntactic Density Score (SDS) With Some Aids for Tabulating," 1973, (ED 091 741).

________, and Carole Kidder, "Syntactic Density and the Computer," Elementary English, vol. 51 (November, 1974), pp. 1128-31.

Greene, Guy S., "The Correlation Between Skill in Performances and Knowledge of Principles in a Course of Speech-Making," Journal of Applied Psychology, vol. 25 (1941), pp. 232-42.

Greene, Harry A., "English Language Grammar and Composition," in Walter S. Monroe (ed.), Encyclopedia of Educational Research, revised edition, New York, Macmillan, 1950, pp. 383-96.

Hall, William E., and James R. Cushing, "The Relative Value of Three Methods of Presenting Learning Material," Journal of Psychology, vol. 24 (July, 1947), pp. 57-62.

Harris, Mary M., "Second Grade Syntax Attainment and Reading Achievement," paper presented at the annual meeting of the International Reading Association, New York, New York, May 1975, (ED 106 764).

Harris, Roland J., "An Experimental Inquiry into the Functions and Value of Formal Grammar in the Teaching of English, with Special Reference to the Teaching of Correct Written English to Children Aged Twelve to Fourteen," unpublished doctor's dissertation, University of London, 1962, in Richard Braddock, Richard Lloyd-Jones, and Lowell Schoer, Research in Written Composition, Champaign, Illinois, National Council of Teachers of English, 1963, pp. 70-83.

Harris, Theodore L., "Reading," in Robert L. Ebel (ed.), Encyclopedia of Educational Research, Toronto, Collier-Macmillan Canada, 1969, pp. 1068-1104.

Heider, Fritz, and Grace Heider, "A Comparison of Sentence Structure of Deaf and Hearing Children," Psychological Monographs, vol. 52 (1940), pp. 42-103.

Heilman, Arthur W., "An Investigation in Measuring and Improving Listening Ability of College Freshmen," Speech Monographs, vol. 28 (November, 1951), pp. 302-08.

Hennings, Dorothy G., "Literature, Language, and Expression," paper presented at the annual meeting of the International Reading Association, Anaheim, California, May, 1976, (ED 124 909).

Heys, Frank, Jr., "The Theme-a-week Assumption: A Report of an Experiment," English Journal, vol. 51 (May, 1962), pp. 320-22.

Hildreth, Gertrude H., "Interrelationships between Written Expression and the Other Language Arts," Elementary English, vol. 31 (January, 1954), pp. 40-48.

________, "Reading Achievement and Linguistic Ability," Education, vol. 69 (May, 1949), pp. 567-72.

Hopkins, Carol J., "An Investigation of the Relationship of Selected Oral Language Measures and First Grade Reading Achievement," Purdue University Report, 1976, (ED 141 782).

Horowitz, Milton W., and Alan Berkowitz, "Listening and Reading, Speaking and Writing: An Experimental

Investigation of Differential Acquisition and Reproduction of Memory," Perceptual and Motor Skills, vol. 24 (February, 1967), pp. 207-15.

Hoyt, Franklin S., "The Place of Grammar in the Elementary Curriculum," Teachers College Record, vol. 7 (November, 1906), pp. 467-500.

Hughes, Richard E., and P.A. Duhamel, Rhetoric: Principles and Usage, 2nd edition, New Jersey, Prentice-Hall, 1967.

Hughes, Theone O., "Sentence Combining: A Means of Increasing Reading Comprehension," 1975, (ED 112 421).

Hughes, Vergil A., "A Study of the Interrelationships Among Selected Language Arts Areas and Abilities," unpublished doctor's dissertation, University of Missouri, 1950, in Dissertation Abstracts International, vol. 10 (1950), p. 65.

Hunt, Kellogg W., Grammatical Structures Written at Three Grade Levels, National Council of Teachers of English Research Report No. 3, Urbana, Illinois, National Council of Teachers of English, 1965.

________, Syntactic Maturity in Schoolchildren and Adults, Monographs of the Society of Research in Child Development, vol. 35, 1970.

Hyndman, Roger, "Some Factors Related to Writing Performance of Tenth-Grade Students," unpublished doctor's dissertation, University of California at Los Angeles, 1969, in Dissertation Abstracts International, vol. 30 (1969), p. 1468A.

Improving Reading Through an Oral Language Program, Bulletin No. 39, Queensland Department of Education, Brisbane, Australia, 1972, (ED 113 701).

Johnson, Gwen P., "An Investigation of Growth in Critical-Reading-Listening Ability in Grades Two and Three," unpublished doctor's dissertation, Florida State University, 1974, in Dissertation Abstracts International, vol. 35 (1974), p. 5779A.

Johnson, Norma, "A Comparison of Syntactic Writing Maturity with Reading Comprehension," paper presented at the annual meeting of the International Reading Association, Texas State Council Conference, 1976, (ED 141 794).

Keltner, John, "The Hardest Knife," Speech Teacher, vol. 6 (November, 1957), pp. 275-84.

Kincaid, Gerald L., "Some Factors Affecting Variations in the Quality of Students' Writing," unpublished doctor's dissertation, Michigan State University, 1953, in Richard Braddock, Richard Lloyd-Jones, and Lowell Schoer, Research in Written Composition, Champaign, Illinois, National Council of Teachers of English, 1963, pp. 83-95.

Kirkham, Sally G., "A Comparison of Silent Reading, Oral Reading, and Listening Comprehension of Students in Grades 3, 6, 9, and 12," unpublished doctor's dissertation, University of Wisconsin-Madison, 1977, in Dissertation Abstracts International, vol. 38 (1977), p. 4458A.

Knower, Franklin H., in John Keltner, "The Hardest Knife," Speech Teacher, vol. 6 (November, 1957), pp. 275-84.

Kuntz, Mildred H., "The Relationship Between Written Syntactic Attainment and Reading Ability in Seventh Grade," unpublished doctor's dissertation, University of Pittsburgh, 1975.

Lacampagne, Robert J., "A National Study of Selected Attitudes and Approaches to Writing of Twelfth-Grade Students with Superior Writing Performance versus those with Average Writing Performance," unpublished doctor's dissertation, University of Illinois, 1968.

Larsen, Robert, and D.D. Feder, "Common and Differential Factors in Reading and Hearing Comprehension," Journal of Educational Psychology, vol. 31 (April, 1940), pp. 241-52.

Lazdowski, Walter P., "Determining Reading Grade Levels from Analysis of Written Compositions," unpublished doctor's dissertation, New Mexico State University, 1976, in Dissertation Abstracts International, vol. 37 (1976), p. 1504A.

Levey, Louise A., "Selected Listening-Reading Skills Relationships in Achieving and Low Achieving Readers of Grade Four," unpublished doctor's dissertation, University of Virginia, 1975, in Dissertation Abstracts International, vol. 36 (1975), p. 4431A.

Levine, Shirley S., "The Effect of Transformational Sentence-Combining Exercises on the Reading Comprehension and Written Composition of Third-Grade Children," unpublished doctor's dissertation, Hofstra University, 1976, in Dissertation Abstracts International, vol. 37 (1976), pp. 6431-32A.

Lewis, Susan E., "Listening Effects Reading Comprehension of Black Inner City Children at Levels One, Two, and Three," unpublished doctor's dissertation, St. Louis University,

1976, in Dissertation Abstracts International, vol. 37 (1976), p. 1962A.

Loban, Walter D., Language Ability: Grades Seven, Eight, and Nine, Washington, U.S. Government Printing Office, 1966.

________, Language Ability: Grades Ten, Eleven, and Twelve, Washington, U.S. Government Printing Office, 1967.

________, The Language of Elementary School Children, NCTE Research Report No. 1, Urbana, Illinois, National Council of Teachers of English, 1963.

Lundstein, Sara W. (ed.), Help for the Teacher of Written Composition (K-9): New Directions in Research, Urbana, Illinois, ERIC Clearing House on Reading and Communication Skills, 1976.

Magnuson, Ralph W., "The Effects of an Alphabet Having a High Sound-Letter Correspondence on Children's Ability to Express Themselves in Written Form," unpublished doctor's dissertation, University of California at Berkeley, 1968, in Dissertation Abstracts International, vol. 30 (1968), p. 1076A.

Maloney, H.B., "An Identification of Excellence in Expository Composition Performance in a Selected 9A Population with an Analysis of Reasons for Superior Performance," unpublished doctor's dissertation, Columbia University, 1967, in Dissertation Abstracts International, vol. 29 (1967), p. 2432.

Martin, Clyde, "Developmental Interrelationships Among Language Variables in Children of the First Grade," Elementary English, vol. 32 (March, 1955), pp. 167-71.

Martin, Elinor B.H., "Administration of the S.O.S. Reading Technique," Edmonton, Alberta, 1976, (duplicated).

Martin, J.J., "The Development of Sentence-Writing Skills at Grades Three, Four and Five," unpublished doctor's dissertation, University of California at Berkeley, 1968, in Dissertation Abstracts International, vol. 30 (1968), p. 1077A.

Martin, P., "A Comparative Analysis of Reading and Writing Skill: Six Case Studies," English in Australia, No. 40, (May, 1977), pp. 51-53.

Martin, R.G., "The Case of Bruce: A First Report on a Secondary Reading Device," Alberta English, vol. 14 (Spring, 1974), pp. 17-21.

________, "Improving Reading Comprehension," a report to the Edmonton Public School Board, June, 1975a, (duplicated).

________, "More Basic Than Basics," Arizona English Bulletin, vol. 20 (October, 1977), pp. 28-33.

________, "A Prediction Formula for a Sample of 'Good' Writing," unpublished doctor's dissertation, University of South Carolina, 1968.

________, "S.O.S.: A Second Report on a Secondary Reading Device," Alberta English, vol. 15 (Winter, 1975-76b), pp. 42-4.

________, "S.O.S. Reading Technique Data Summary from McDougall Junior High School Davis Test Score Differences, 1974-75," a report to the Edmonton Public School Board, 1975c, (duplicated).

________, and J.F. Belanger, "A Longitudinal Assessment of Effects of S.O.S. on Reading and Writing at Two Grade Levels," final report to the Edmonton Public School Board, March, 1977, (duplicated).

________, and J.F. Belanger, Project Proposal, "A Longitudinal Assessment of Effects of S.O.S. on Reading and Writing at Two Grade Levels," request to the Edmonton Public School Board, October, 1975, (duplicated).

________, and Elinor B.H. Martin, "S.O.S.: A Forty-Minute Technique," Edmonton, Alberta, 1974, (multilithed).

Mathews, Ernest G., Robert P. Larsen, and Gibbon Butler, "Experimental Investigation of the Relation Between Reading Training and Achievement in College Composition Classes," Journal of Educational Research, vol. 38 (March, 1945), pp. 499-505.

Matt, David W., "An Enquiry into Empirical Relationships Between the Reading and Writing of Exposition and Argument," unpublished doctor's dissertation, State University of New York at Albany, 1977, in Dissertation Abstracts International, vol. 38 (1977), pp. 4631-32A.

Mazurkiewicz, Albert J., "A Comparison of i.t.a. and T.O. Reading, Writing, and Spelling Achievement when Methodology is Controlled," in I.T.A. and the World of English, the I.T.A. Foundation, Hofstra University, Hampstead, New York, 1966.

________, "Comparative Attitudes and Achievements of the 1963 i.t.a. and T.O. Taught Students in the Tenth and Eleventh Grades," Reading World, vol. 15 (May, 1975),

pp. 242-51.

________, "The Early to Read-i.t.a. Program: Effects and Aftermath of a Six Year Longitudinal Study," 1971, (ED 059 850).

________, "The Initial Teaching Alphabet in Reading Instruction," Lehigh-Bethlehem Evaluation: Demonstration Project on the Use of i.t.a., comprehensive final report, February, 1967, (ED 063 082).

________, "i.t.a. Revisited," paper read to the College Reading Association, 1973, (ED 083 569).

McCaffrey, Judith P., "Grammar and Reading: A Study of Some Relationships Between Selected Sentence Structures and Reading Comprehension for Students in Grades 2, 3, and 4," unpublished doctor's dissertation, Harvard University, 1976, in Dissertation Abstracts International, vol. 37 (1976), p. 4117A.

McCrery, Lester L., "Relationship Between Writing and Speaking as Measured by College Grades and Student Rating Scales," Journal of Communication, vol. 1 (May, 1951), pp. 40-44.

Meckel, Henry C., "Research on Teaching Composition and Literature," in N.L. Gage (ed.), Handbook of Research on Teaching, Chicago, Rand-McNally, 1963, pp. 966-90.

Mellon, John C., Transformational Sentence Combining, A Method for Enhancing the Development of Syntactic Fluency in English Composition, NCTE Research Report No. 10, Champaign, Illinois, National Council of Teachers of English, 1969.

Miller, Tyree J.A., "A Quantitative Study of the 'Free Modifiers' in Narrative-Descriptive Compositions Written by Black College Freshmen after Leaving the Influence of the Christensen Program and a Study of their Attitudes Toward Written Composition," unpublished doctor's dissertation, Ohio State University, 1974, in Dissertation Abstracts International, vol. 33 (1974), p. 3483.

Miller, B.D., and J.W. Ney, "The Effect of Systematic Oral Exercises on the Writing of Fourth-grade Students," Research in the Teaching of English, vol. 2 (1968), pp. 44-61.

Mills, Editha B., "An Experimental Study in the Use of Literary Models in Written Composition," unpublished doctor's dissertation, University of Georgia, 1967.

Monk, Richard J., "A Study to Determine the Relationship

Between Children's Home Environments and their School Achievement in Written English," unpublished doctor's dissertation, University of Washington, 1958.

Moore, Paul L.A., "A Descriptive Analysis of R. Zoellner's Talk-Write Theory: A Behavioral Pedagogy for Composition," unpublished doctor's dissertation, State University of New York at Buffalo, 1971, in Dissertation Abstracts International, vol. 32 (1971), p. 2534.

Nagle, John E., "The Effects of a Directed Writing Activity in Eighth Grade Social Studies Instruction on General Reading Achievement and on Social Studies Reading Achievement," unpublished doctor's dissertation, Temple University, 1972, in Dissertation Abstracts International, vol. 33 (1972), p. 1523A.

Nichols, Ralph G., "Factors in Listening Comprehension," Speech Monographs, vol. 15 (1948), pp. 154-63.

________, and Robert J. Keller, "The Measurement of Communication Skills," Junior College Journal, vol. 24 (November, 1953), pp. 160-68.

Nikos, G.B., "ITA and TO - Their Impact on Spelling and Writing," The Elementary School Journal, vol. 70 (1970), pp. 321-30.

Obenchain, Anne, "Effectiveness of the Precis Essay Question in Programming the Sequential Development of Written Composition Skills and the Simultaneous Development of Critical Reading Skills," unpublished master's thesis, George Washington University, 1971, in Sandra L. Stotsky, "Sentence-Combining as a Curricular Activity: Its Effect on Written Language Development and Reading Comprehension," Research in the Teaching of English, vol. 9 (Spring, 1975), pp. 30-71.

O'Dea, Paul, "Five Myths in the Teaching of Composition," English Journal, vol. 54 (April, 1965), pp. 328-30.

O'Donnell, John F., "An Experimental Study of the Effects of the Supplemental Use of a Psycholinguistic Remedial Tutorial Program on the Reading and Writing Behaviors of Black, High-Risk College Freshmen, and on their Attitudes Toward Reading, Writing and Other College-Related Stimuli," unpublished doctor's dissertation, Temple University, 1974, in Dissertation Abstracts International, vol. 35 (1974), p. 1552A.

O'Donnell, Roy C., "The Correlation of Awareness of Structural Relationships in English and Ability in Written Composition," Journal of Educational Research, vol. 57 (May, 1964), pp. 464-67.

________, "A Critique of Some Indices of Syntactic Maturity," Research in the Teaching of English, vol. 10 (Spring, 1976), pp. 31-38.

Oehlkers, William, "The Contribution of Creative Writing to Reading Achievement in the Language Experience Approach," unpublished doctor's dissertation, University of Iowa, 1972.

O'Hare, Frank, Sentence Combining: Improving Student Writing Without Formal Grammar Instruction, NCTE Research Report No. 15, Urbana, Illinois, National Council of Teachers of English, 1973.

Olson, Patricia H., "Relationship of Selected Oral Language Components to Reading Achievement of Third Grade Students," unpublished doctor's dissertation, University of Arizona, 1976, in Dissertation Abstracts International, vol. 37 (1976), p. 1390A.

Palmer, William S., "Reading, Writing and the Realm of Reason," Phi Delta Kappan, vol. 52 (April, 1971), pp. 473-75.

Parke, Margaret B., "Composition in the Primary Grades," in Alvina T. Burrows (ed.), Margaret B, Parke, Neal R. Edmund, John J. DeBoer, Thomas D. Horn, Virgil E. Herrick, and Ruth C. Strickland, Children's Writing: Research in Composition and Related Skills, Champaign, Illinois, National Council of Teachers of English, 1961, pp. 6-20.

________, "When They are Six," unpublished doctor's dissertation, Columbia University, 1945, in Burrows, et al., Children's Writing: Research in Writing and Related Skills, Champaign, Illinois, NCTE, 1961, p. 17.

Perrin, Porter G., and Wilma R. Ebbitt, Writer's Guide and Index to English, 5th edition, Glenview, Illinois, Scott-Foresman, 1972.

Perron, J.D., "The Impact of Mode on Written Syntactic Complexity: Part III Fifth Grade Report," October, 1976, (ED 126 511 and 126 531).

Piexotto, Helen E., "The Relationship of College Board Examination Scores and Reading Scores for College Freshmen," Journal of Applied Psychology, vol. 30 (August, 1946), pp. 406-11.

Potter, Robert R., "An Exploratory Study of the Relationship Between Certain Aspects of Sentence Structure and the Overall Quality of Tenth-Grade Writing," unpublished

doctor's dissertation, Teachers College, Columbia University, 1966.

Quinn, Anne-Marie V., "Some Effects of Experience-Related Reading Instruction on the Written Language Development of First-Grade Children," unpublished doctor's dissertation, University of Maine, 1977, in Dissertation Abstracts International, vol. 38 (1977), p. 1946A.

Radcliffe, Terrence R., "An Experimental Investigation to Evaluate the Effectiveness of Talk-Write Composition," unpublished seminar report, University of Montana, 1969a.

________, "The Results of an Empirical Study Using Talk-Write Composition," unpublished pilot study, University of Montana, 1969b, (mimeographed).

________, "Talk-Write Composition: An Instructional Method Utilizing Speech to Improve Writing Ability," unpublished paper, University of Montana, 1971.

________, "Talk-Write Composition: A Method for the Teaching of Writing," unpublished booklet, College of Great Falls, 1969c.

________, "Talk-Write Composition: A Theoretical Model Proposing the Use of Speech to Improve Writing," Research in the Teaching of English, vol. 6 (Fall, 1972), pp. 187-99.

Reed, Estelle, "Improving Comprehension Through Study of Syntax and Paragraph Structure in Seventh Grade Classes," in J.A. Figurel (ed.), Forging Ahead in Reading, Proceedings of the International Reading Association, vol. 12 (1967), pp. 575-79.

Rice, J.M., "English, The Need for a New Basis in Education," The Forum, vol. 35 (January, 1903), p. 451.

Rinne, C.H., "Improvement in Reading Comprehension Through Increased Awareness of Written Syntactic Patterns," unpublished doctor's dissertation, Stanford University, 1967.

Robertson, Jean E., "An Investigation of Pupil Understanding of Connectives in Reading," unpublished doctor's dissertation, University of Alberta, 1966.

Rodgers, D., K. Slade, and R. Conry, "Oral Language Reading Ability, and Socioeconomic Background in Three Grade One Classes," Alberta Journal of Educational Research, vol. 20 (December, 1974), pp. 316-26.

Rosenbaum, Peter S., in Frank O'Hare, Sentence Combining: Improving Student Writing without Formal Grammar Instruction, NCTE Research Report No. 15, Urbana, Illinois, National Council of Teachers of English, 1973, pp. v-vi.

Rossignol, Lois J., The Relations Among Hearing Acuity, Speech Production, and Reading Performance in Grades 1A, 1B, and 2A, Teachers College Contributions to Education, No. 936, New York, Teachers College, Columbia University, 1948.

Ruddell, Robert B., "The Effect of Oral and Written Patterns of Language Structures on Reading Comprehension," The Reading Teacher, vol. 18 (January, 1965), pp. 270-75.

Russell, David H., and Henry R. Fea, "Research on Teaching Reading," in N.L. Gage (ed.), Handbook of Research on Teaching, Chicago, Rand-McNally, 1963, pp. 825-928.

Sandel, Lenore, "A Comparison Between the Oral and Written Responses of First Grade Children in ITA and TO Classes," 1967, (ED 019 144).

________, "A Comparison of Oral and Written Language of Third-Grade Children Instructed in i.t.a. and TO in First Grade," unpublished doctor's dissertation, Hofstra University, 1970.

Schneider, Virginia L., "A Study of the Effectiveness of Emphasizing the Teaching of Reading Skills to Improve Composition Skills in Remedial English Classes at Kansas City Community Junior College," unpublished doctor's dissertation, University of Kansas, 1971.

Schonell, Fred J., Backwardness in the Basic Subjects, Toronto, Clarke, Irwin, 1942.

Schulwitz, Bonnie S., "An Investigation of the Effect of Listening Instruction Upon the Reading Comprehension of First Grade Pupils," unpublished doctor's dissertation, Michigan State University, 1973, in Dissertation Abstracts International, vol. 34 (1973), p. 5496A.

Searles, John R., and G. Robert Carlsen, "English," in Chester W. Harris (ed.), Encyclopedia of Educational Research, third edition, New York, Macmillan, 1960, pp. 454-70.

Secondary Language Arts Handbook Grades 7-12, Edmonton, Alberta, Queens Printer, 1972.

Selover, Robert B., and James P. Porter, "Prediction of the Scholarship of Freshman Men by Tests of Listening and

Learning Ability," Journal of Applied Psychology, vol. 21 (1937), pp. 583-88.

Settles, Ivan L., "The Effects of Instruction with the Initial Teaching Alphabet on the Written Composition of Primary Students," unpublished doctor's dissertation, Washington State University, 1967, in Dissertation Abstracts International, vol. 28 (1967), p. 2459.

Shackford, Helen G., "Junior High School Student's Knowledge of Grammatical Structure and its Relation to Reading Comprehension," unpublished doctor's dissertation, Boston University, School of Education, 1976, in Dissertation Abstracts International, vol. 37 (1976), p. 1380A.

Shapiro, Bernard J., and Phyllis P. Shapiro, "The Effect of Reading Method on Composition: ITA vs. TO," Journal of Reading Behavior, vol. 5 (Spring, 1973), pp. 82-87.

________, "The Relationship Between Reading Medium and Prose Writing," paper presented to the American Educational Research Association, 1973, (ED 082 145).

Sheils, Merrill, "Why Johnny Can't Write," Newsweek, vol. 86 (December 8, 1975), pp. 58-65.

Sherwin, J.S., Four Problems in Teaching English: A Critique of Research, Scranton, Pennsylvania, International Textbook, 1969.

Shire, Sister Mary Louise, "The Relation of Certain Linguistic Factors to Reading Achievement of First Grade Children," unpublished doctor's dissertation, Fordham University, 1945.

Siedow, Mary D., "Relationships Between Syntactic Maturity in Oral and Written Language and Reading Comprehension of Materials of Varying Syntactic Complexity," unpublished doctor's dissertation, Indiana University, 1973.

Smith, Carl B., "Relation of First Grade Reading and Composition," paper presented to the International Reading Association, 1968, (ED 020 080).

Squire, James R., and Roger K. Applebee, High School English Instruction Today, The National Study of High School English Programs, New York, Appleton-Century-Crofts, 1968.

Steinberg, Edwin R., Needed Research in the Teaching of English, Cooperative Research Monograph No. 11, Washington, U.S. Government Printing Office, 1963.

Stern, A.A., "Semantic Analysis: A Reading Approach to the Teaching of Composition," unpublished doctor's dissertation, Columbia University, 1972, in Dissertation Abstracts International, vol. 33 (1972), p. 226A.

Sternglass, Marilyn S., "Composition Teacher as Reading Teacher: Linguistics is the Bridge," paper presented at the annual meeting of College Composition and Communication, Philadelphia, March, 1976, (ED 120 782).

Stewart, R.W., "Curriculum Change as the Result of Introducing Children to Reading Through ita in Bethlehem," The ita Journal, (June, 1969), pp. 17-22.

Stewig, John W., Read to Write: Using Children's Literature as a Springboard to Writing, New York, Hawthorne, 1975, (ED 136 274).

Stoodt, Barbara, "The Relationship Between Understanding Grammatical Conjunctions and Reading Comprehension," unpublished doctor's dissertation, Ohio State University, 1970, in Dissertation Abstracts International, vol. 31 (1970), p. 4635A.

Stotsky, Sandra L., "Sentence-Combining as a Curricular Activity: Its Effect on Written Language Development and Reading Comprehension," Research in the Teaching of English, vol. 9 (Spring, 1975), pp. 30-71.

Strickland, Ruth G., "Evaluating Children's Compositions," Elementary English, vol. 37 (May, 1960), pp. 321-30.

________, The Language of Elementary School Children: Its Relationship to the Language of Reading Textbooks and the Quality of Reading of Selected Children, Bulletin of the School of Education, Bloomington, Indiana, Indiana University, vol. 38, 1962.

Sullivan, Edmund V., Moral Learning, Findings, Issues, and Questions, Toronto, Paulist Press, 1975.

Swain, Louis L., "Does Public Speaking Teach Written Usage?" Quarterly Journal of Speech, vol. 36 (April, 1950), pp. 220-25.

Taylor, Nancy E., "An Investigation of Children's Awareness of the Acceptability of Selected Grammatical and Semantic Structures and Their Relationship to Reading Achievement, unpublished doctor's dissertation, University of Maryland, 1976, in Dissertation Abstracts International, vol. 37 (1976), p. 4190A.

Thomas, Freddy L., "The Extent of the Relationship Between Reading Achievement and Writing Achievement Among College

Freshmen," unpublished doctor's dissertation, University of South Carolina, 1976, in Dissertation Abstracts International, vol. 37 (1976), p. 6320A.

Thompson, Wayne N., Quantitative Research in Public Address and Communication, New York, Random House, 1967.

Tovatt, Anthony L., and Ebert L. Miller, "Oral-Aural Visual Stimuli Approach to Teaching Written Composition to 9th Grade Students," Report No. BR-5-0398, Muncie, Indiana, Burris Laboratory School, Ball State University, January, 1967, (ED 015 204).

Trost, David M., "Attitudes, Writing Fluency, Reading Achievement - A Comparison Between ITA and TO Trained Children," unpublished doctor's dissertation, University of Michigan, 1971, in Dissertation Abstracts International, vol. 32 (1971), p. 6114.

Weaver, Phyllis A., "Component Skills in Reading Comprehension: An Investigation of the Effects of Sentence Organization Instruction," paper presented at the annual meeting of the American Educational Research Association, April, 1977, (ED 136 241).

Weidner, Margaret J., "A Study of the Effects of Teacher Oral Reading of Children's Literature on the Listening and Reading Achievement of Grade Four Students," unpublished doctor's dissertation, Boston University, School of Education, 1976, in Dissertation Abstracts International, vol. 37 (1976), p. 4849A.

Weinstein, R., and M.S. Rabinovitch, "Sentence Structure and Retention in Good and Poor Readers," Journal of Educational Psychology, vol. 62 (1971), pp. 25-30.

Whittaker, Jeweleanne W., "Effects of the Application of Linguistics on Reading Comprehension of Black Freshmen Students at Texas Southern University, unpublished doctor's dissertation, University of Houston, 1974, in Dissertation Abstracts International, vol. 35 (1974), pp. 1886-87A.

Wiggins, Robert F., "A Study of the Influence of Oral Instruction of Students' Ability in Written Sentence Structure," unpublished doctor's dissertation, University of South Carolina, 1968, in Dissertation Abstracts International, vol. 29 (1968), pp. 3927-28A.

Wilkinson, Andrew, "Oral Constraints and Reading Acquisition," Educational Review, vol. 22 (1969), pp. 103-16.

Winter, Blotilda, "Interrelationships Among Variables in

Children of the First and Second Grades," Elementary English, vol. 34 (February, 1957), pp. 108-13.

Winterowd, W. Ross, "Style: A Matter of Manner," Quarterly Journal of Speech, vol. 56 (1970), pp. 161-67.

Wiseman, Stephen, "The Marking of English Composition in Grammar School Selection," British Journal of Educational Psychology, vol. 19 (November, 1949), pp. 200-09.

________, "Reliability and Validity," British Journal of Educational Psychology, vol. 26 (November, 1956), pp. 172-79.

Wixon, Vincent, and Patty Stone, "Getting it Out, Getting it Down: Adapting Zoellner's Talk-Write," paper presented to the National Council of Teachers of English Secondary School English Conference, Seattle, Washington, March, 1977, (ED 140 354).

Wolfe, Rosemary F., "An Examination of the Effects of Teaching a Reading Vocabulary upon Writing Vocabulary in Student Compositions," unpublished doctor's dissertation, University of Maryland, 1975, in Dissertation Abstracts International, vol. 36 (1975), pp. 3337-38A.

Wood, Donna S., "The Relationship Between Rate Controlled Speech and Reading Comprehension Skills of Postsecondary Technical Students," unpublished doctor's dissertation, Memphis State University, 1975, in Dissertation Abstracts International, vol. 36 (1975), p. 5797A.

Woodward, John C., and Arthur G. Phillips, "Profile of the Poor Writer - The Relationship of Selected Characteristics to Poor Writing in College," Research in the Teaching of English, vol. 1 (Spring, 1967), pp. 41-53.

Wyatt, Nita, "A Study of the Relationship of Extensive Reading to Certain Writing Skills of a Selected Group of Sixth Grade Children," unpublished doctor's dissertation, University of Kansas, 1960, in Dissertation Abstracts International, vol. 21 (1960), p. 2643.

Wykoff, George S., "Reading as a Problem in Composition," College English, vol. 4 (January, 1943), pp. 245-54.

Young, William E., "The Relation of Reading Comprehension to Hearing Comprehension and Retention," Journal of Experimental Education, vol. 5 (1936), pp. 30-39.

Zanotti, Robert J., "A Study of the Use of the Tape Recorder as an Aid to Written Composition at the Sixth-Grade

Level," unpublished doctor's dissertation, State University of New York at Buffalo, 1970, in Dissertation Abstracts International, vol. 31 (1970), p. 1520.

Zeman, Samuel S., "Reading Comprehension and the Writing of 2nd and 3rd Graders," Reading Teacher, vol. 23 (November, 1969), pp. 144-50.

________, "The Relationship Between the Measured Reading Comprehension and the Basic Sentence Types and Structural Patterns in Composition Writing by Second and Third Grade Children," unpublished doctor's dissertation, Lehigh University, 1966.

Zoellner, Robert, "Talk-Write: A Behavioral Pedagogy for Composition," College English, vol. 30 (January, 1969), pp. 267-320.

APPENDIX A

The S.O.S. Reading Technique

Two previous experiments, experiments of a smaller scale, indicated that the S.O.S. Reading Technique was an effective vehicle for enhancing students' reading abilities. The current experiment, as noted in Chapter I, was designed as the writing portion of a study entitled "A Longitudinal Assessment of Effects of S.O.S. on Reading and Writing at Two Grade Levels" (Martin and Belanger, 1977).

1. Description and Strategy of the S.O.S. Reading Technique. Materials for the S.O.S. consist of one slip of paper and four printed sheets. The slip of paper contains five columns of very simple words already familiar to the student and involving only the "short" vowels (e.g., a list of bag, nap, hat) such that a student can pick out the common sound in each group of words without being told and thereby establish the vowels that will be used throughout the remaining pages; the slip of paper also clarifies how ch will be used, so that all spellings will be phonologically unambiguous. The first of the four printed sheets contains only nonsense syllables; the second contains nonsense syllables and words; the third and fourth contain words only. Administration of the technique requires that each subject, up to four in number, read the lines and columns of print, individually, in turn, and aloud, while the others follow along silently.

Page one has 28 horizontal lines of nonsense syllables, graduated in length and difficulty down the page. To read these aloud correctly, the student is forced to rely

on the sounds represented by the letters. The administrator halts him on any error but without supplying the correct sound. Page two, a transition page, has alternate vertical columns of nonsense syllables and words. Each subject in turn reads a column of ten nonsense syllables and then a column of ten words, reading the words somewhat more rapidly than the nonsense syllables. Sheets three and four each consist of ten columns of twenty words. Subjects read the columns aloud and in turn, fast enough to lose conscious control.

The four pages of the technique require the subject to rely on his knowledge of the sound system of written English, a knowledge, it is hypothesized, gained from years of exposure to the printed page (Martin and Martin, 1974, pp. 16-20). The strategy of the technique, as explained by its inventor, is as follows:

> The technique is designed to <u>subtract</u> from the student his present burden of word ritual by forcing him to rely on sound for the nonsense syllables on Sheet 1, then to enable him to transfer his meaningless reading from nonsense syllables to words on Sheet 2, then to cause him to lose conscious control of meaningless words on Sheets 3 and 4. The aim is to enable him to hear print vicariously by taking his focus of attention off each individual word. The meaning will then fall on the sentence as it does in hearing. (E. Martin, 1976).

A complete description and rationale of the technique (including, e.g., why "long" vowels are omitted) and its theoretical basis as well as instructions for its use are found in Martin and Martin (1974), <u>S.O.S.</u>: <u>A Forty-Minute Reading Technique</u>, and E. Martin (1976), <u>Administration of</u>

<u>the S.O.S. Reading Technique</u>. A more concise account is contained in R. Martin (1977).

2. <u>Experimental Data</u>. In two previous experiments, Martin (1975a, 1975c) reported that students administered the S.O.S. Reading Technique made statistically significant gains in reading ability when compared with students who had not been administered the technique.

a. <u>the Rosslyn experiment</u>. A six-week experiment (Martin, 1975a) was conducted with 56 junior-high-school students randomly assigned to experimental and control groups. The treatment was administered to the experimental group by the students' classroom teacher during the lunch hour and after school over a ten-day period. Pre-post gains were 3.63 points for the experimental group and .40 points for the control group, statistically significant beyond the .05 level of confidence.

Martin noted three testing anomalies:

i. Davis Form 2A, the posttest, had been administered to both the experimental and control groups nine months previously.

ii. The Edmonton Public School Board testing service reported the results of the pretest, Davis Form B, in terms of "raw-score correct" figures rather than using the Davis correction formula. The investigator felt that this favored the control group.

iii. A fifty-item "cloze" test designed by the investigator showed slight and non-significant differences in favor of the control group. He attributed this to the stated purpose of the technique: to reduce word guessing.

Experimental students in the Rosslyn experiment were also administered an attitude questionnaire. Analysis of this questionnaire showed that 85 percent of the experimental students reported that they found reading easier, 80 percent reported that they now did less re-reading to get the idea of a passage, 96 percent thought that the technique would be helpful to other students, and 71 percent reported that they were reading faster.

b. the McDougall experiment. Martin (1975c) reported a semi-formal junior-high-school experiment conducted by a classroom teacher. Volunteer subjects from two classes were used in the experiment: in Class A, 16 subjects volunteered to take the treatment while 6 did not: in Class B, 17 volunteered while 3 did not. Class C, taught by another teacher, was used for purposes of comparison. The treatment was administered in late February and was followed by the system-wide administration of the Davis Form 2B in early March. The Davis Form 2A, which all students had taken the previous September, was used again in June as a posttest. Computed on unadjusted raw scores, the three-month mean gain for the control group was 2.1 points and for the experimental

group 6.9 points, a difference statistically significant beyond the .02 level of confidence.

c. <u>the EPSB multi-school experiment</u>. The following pages of Appendix A contain the official report to Edmonton Public Schools for the experiment (Martin and Belanger, 1977) associated with the one reported in this thesis. It should be noted that for reasons discussed in footnote 1, Chapter IV, the n of 205 differs slightly from that of the present experiment.

March 28, 1977

Dr. Tom Blowers
Director of Research
Edmonton Public Schools
10010 - 107A Avenue
Edmonton, Alberta T5H 0Z8

Dear Tom:

Attached is the "Research Summary" for the project entitled "A Longitudinal Assessment of Effects of S.O.S. on Reading and Writing at Two Grade Levels", as originally approved by you on November 5, 1975.

In addition, as discussed in our recent phone conversation and my letter of March 16, I will forward you any further material concerning this project, and Mr. Joe Belanger will send a copy of his thesis on the writing portion when it is completed. The two-page "Data Sheet" attached as an appendix to the report contains a reasonably full summary of the reading data by grade, class, and sex. I will also be glad to let you look at the computer printouts, answer sheets on the Davis testing, etc. if you would like a still fuller picture of the reading portion of the project. Mr. Belanger's thesis will provide abundant material supplemental to the present rather compact summary of the writing portion of the project, and he will be glad to furnish or lend any of the "physical" data in which you may be interested.

My thanks are extended to you and to all in the EPSB system who cooperated to make this project possible.

Sincerely,

R. Glenn Martin
Associate Professor
Secondary Education (English)

RESEARCH SUMMARY

Title of Research Project A Longitudinal Assessment of Effects of S.O.S. on Reading and Writing at Two Grade Levels

Name(s) of Researcher(s) R. Glenn Martin and Joseph F. Belanger

University Department Secondary Education

M.Ed. Thesis [ ] Ph.D. Thesis [X] Project [X] Other (specify) Writing: Ph. D. Thesis Reading: Project

SPECIFIC PROBLEM AREA EXAMINED

This research was concerned with the improvement of reading and writing ability at secondary level. It was a sequel to previously reported research at Rosslyn and McDougall Junior High Schools (1974-75) on the effectiveness of a one-time, 40-minute use of the "S.O.S. Reading Technique" in improving reading achievement at ninth-grade level. These earlier studies, both small, had shown statistically significant gains for S.O.S. experimentals over controls. The present research sought, in the first instance, to confirm or otherwise the findings of the earlier studies. It aimed additionally, with a broader sample, to see whether use of S.O.S. would yield differential results if applied to both ninth- and tenth-grade students and to determine whether, if initial gains were achieved after three months, they would be sustained after a six-month period. Concurrently, an estimate was sought as to any effects on writing that might appear if reading results had been achieved. Pilot work, clinical data, and the familiar correlations between reading and writing proficiency lent plausibility to this second main question.

MAJOR HYPOTHESES OR OBJECTIVES OF STUDY

The study sought to determine:

(1) Effectiveness of S.O.S. on reading achievement as measured by Davis Series 2 Reading Tests over a larger sample than those previously examined.

(2) Differential effectiveness, if any, between use at Grade IX and use at Grade X.

(3) Effectiveness as measured three months and six months after administration.

(4) Answers to identical questions with respect to writing proficiency as measured by a general subjective assessment of quality, by raters, and three quantitative analyses of writing samples taken at the same intervals as in the case of reading.

DESCRIPTION OF THE SAMPLE

Four pairs of classes were used, each pair taught by the same teacher and undergoing the same instruction, two pairs at ninth-grade level and two pairs at tenth-grade level. Classes were selected in consultation with central office personnel to be reasonably typical of the system and to have no direct involvement in other experimental procedures in reading or in composition during the six-month posttest period. At Grade X, only English 10 classes were used (and not English 13) for logistical reasons within the school. Combined 8-class enrollment was 230. Students were excluded from data analysis if they missed any of the three reading tests; this left an N of 205 for the reading analysis. An additional 11 students, who missed either reading or writing tests, were excluded from the writing analysis, leaving a writing N of 194. Certain reading-writing comparisons for the 194-student sample will be reported in the Belanger thesis.

BRIEF SUMMARY OF RESEARCH PROCEDURES Of each pair of classes (described in "Sample"), one was randomly assigned to experimental (treatment) group and one to control (non-treatment) group. No placebo was considered necessary because of the short (40-minute), single treatment. Design involved, for both reading and writing, a pretest, treatment, three-month posttest and six-month posttest, with analysis of data by ANCOVA-35 (3-Way Analysis of Variance and Covariance) at the Computer Centre of the U. of A. Faculty of Education. Instruments for reading measurement were Forms D, B, and C of the Davis Series 2 Reading Tests, administered by J. Belanger and scored, using Davis correction formula ($R-\frac{1}{4}W$), with assistance from the Testing Centre of EPSB. Treatment was by Elinor Martin in groups of 4 students at a time over a two-day period in the school for each section involved. Writing samples were likewise collected by J. Belanger in each school. Topics were rotated, and detailed analysis of subjective quality estimates and objective measures of syntactic density, T-unit length, and theme length performed for sub-samples of 72 students and 96 students on the first two measures, and for all students on the third and fourth measures. Treatment was in early December, 1975. A more detailed account of the writing procedures will be reported in the Belanger thesis, along with a number of analyses relating reading to writing that are not included in this report.

RESULTS OF THE PROJECT (State major findings in point form)

I. Reading results (see Appendix A, "Data Sheet", for main numerical findings)

1. Reading results confirmed earlier indications (Rosslyn and McDougall experiments in 1974-75) that S.O.S. administration produces significantly better gains in reading competence, as measured by Davis reading tests, among ninth-grade classes than are achieved by comparable control classes.
2. Reading results demonstrated comparable gains for tenth-graders who have had S.O.S. as contrasted with control groups.
3. Gains achieved at the end of a three-month period are sustained at the end of a six-month period. Mean raw score gain after three months (as between experimentals and controls) was 2.904, and the difference after six months was 3.041 raw score units (adjusted for guessing by the Davis correction formula).
4. Differences in experimentals' gains were not significant for grade or sex (or, to be more precise, there was no statistically significant interaction between treatment and grade or treatment and sex), though inspection of the data reveals nonsignificant differences in favor of girls and tenth-graders. No formal test was performed on the effect of initial ability level on results, but again inspection of data suggests somewhat wider gains for students with higher initial scores.
5. Overall differences in gains between experimental and control students were significant at about the .001 level.

II. Writing results (see Appendix A, "Data Sheet," for main numerical findings)

1. On four assessments of writing (subjective estimates of quality, a measure of syntactic density based on a correction of Golub's measure, mean length of T-units [a refined version

OVER

of the cruder sentence length], and on fluency [mean number of words written]) there were no significant differences between controls and experimentals, overall or by grade or sex. Nonsignificant differences favored controls on subjective quality, T-unit length, and fluency, and experimentals on syntactic density, by small margins in all cases.

2. Data thus exhibited no appreciable effect on writing as a result of S.O.S. where normal curriculum is adhered to in the succeeding three-month or six-month periods.

IMPLICATIONS OF THE RESULTS (For classroom, school or school system practice; for development of theory)

Reading implications

1. The results confirm those obtained earlier in experiments at Rosslyn and McDougall Junior High Schools as to the general effectiveness of S.O.S. in relation to the expenditure of time and resources involved. The greater statistical significance of the results in the present experiment is presumably attributable to the larger number of pupils involved, since it is, in fact, somewhat less between any pair of classes in the present experiment than between the two pairs of classes in the earlier work. This result is noteworthy in that in the two former cases, administration was by the students' own classroom teacher, while in the present case it was by a stranger who was, however, more technically proficient in its use. Optimal reading results would appear likely to result from administration by the students' own classroom teacher with somewhat fuller instruction in its use than was provided to the teachers in the Rosslyn and McDougall cases. On the other hand, results in these three experiments are unlikely to be matched by teachers who are less outstanding or less in sympathy with the technique's rationale than in the three classroom experiments to date. Widescale use of the technique, therefore, would involve very considerable attention to the quality of administration, if comparable results are to be achieved. The degree to which the rationale of the technique runs counter to the "received wisdom" about reading looms as an important consideration in any plans for wider-scale use.
2. The data, revealing slightly (nonsignificantly) greater benefits for tenth-graders, girls, and (presumably) higher-ability students, confirm the view that the technique is applicable to the broad spectrum of secondary students (including, from other data, English 30) and not merely, in the ordinary sense, a device for standard "remedial" cases. The present results support the view that if S.O.S. is used, it should be used with entire classes (four students at a time, of course), since it appears to benefit those who are considered normal readers just as much as, or more than, those who are in obvious need of special reading help. This also avoids stigmatizing S.O.S. as "remedial". (An apter expression would be that it is a "comprehension improver".)
3. The implications of sustaining the gain on the second posttest are that S.O.S. is not "instructional", since some loss in "learning" from the fourth to the sixth month after administration ought to be observed if it were. Thus it would appear that what is affected is basic competence in reading, at all ability levels, and not the acquiring of new information or a new isolated "skill", whose retention ought to be fading after six months. This implication, of course, would be further clarified if students in the

CONTINUED ON FOLLOWING PAGE

PLEASE FORWARD THE COMPLETED RESEARCH SUMMARY TO THE DIRECTOR OF EDUCATIONAL RESEARCH, EDMONTON PUBLIC SCHOOL BOARD, 10010-107A AVENUE, EDMONTON

experiment were followed in a way that permitted still later comparisons between them.

4. Although experimental design called for making the technique available to controls following the experiment, no such arrangements in fact materialized (except with one student). While Hawthorne effect might therefore be invoked as an explanation of results, it appears very doubtful that a single experience lasting less than an hour in December would have marked attitude effects on performance in June. Two other considerations further reduce the plausibility of Hawthorne explanations: (1) recent research on the Hawthorne effect itself (Cook in 1967) casts heavy doubt on its long-standing reputation as a serious influence on school results, and (2) if it is supposed that S.O.S. works primarily by producing attitude change in students, one would suppose that this would also be reflected in changes in writing, especially in view of the well-known correlations between reading and writing proficiency (further confirmed for a subsample in the present study); in fact, however, the second phase of the present experiment (on writing) did not demonstrate change. (This is further discussed below.)
5. S.O.S. clearly warrants still further experimentation, preferably not under the control of the sponsors of the technique. The problems of administration, however, are such that those administering it would, at present, have to receive specific instruction from the technique's sponsors, and that not all teachers are, in fact, able to do it competently. This poses logistical problems in outside replication. (One such arrangement has recently been requested by a well-known researcher at the University of Toronto.)
6. The Davis reading tests, although they have generally received top reviews among the available alternatives for measuring reading proficiency, are not considered by these researchers to be optimal, or even especially suitable, for measuring the capacity which S.O.S. apparently affects (cf. current theoretical material by its sponsors, describing the rationale in some detail). Davis results clearly are heavily influenced by both vocabulary range and measurable general intelligence, neither of which could possibly be affected in the short run by S.O.S. The optimal test, not now in

Reading implications (continued)

existence, would measure the student's ability to react to written language as he would have reacted to it had he heard it, and a student scoring 24 on the Davis test might well score higher on such a test than would another student, with larger vocabulary and higher general intelligence, who scored 48 on the Davis test; and in this sense, a student with 24 on Davis might well be a "better reader" than one with 48 or even 72. Present results should be viewed in this context.

Writing implications

1. Earlier pilot testing on writing effects of S.O.S. was totally unconfirmed by the present experiment. The simplest interpretation is that S.O.S. does not affect writing. Attempts back to 1931 (Eurich) had attempted to demonstrate the effect of reading improvement on writing, but all had failed because they could not change reading; the present attempt surmounted this hurdle but could not demonstrate writing change. The simplest explanation fails, however, to account for earlier pilot results and clinical data, some of them quite impressive, and also fails to reflect the well-substantiated correlations between reading and writing proficiency. If these correlations are real (and they are about +.50 for the "quality" subsample of 72 students--.466 for pretest, .535 for first posttest, and .477 for second posttest), it would then appear that a change in reading proficiency should be accompanied by at least some change in writing proficiency. In the present case, it may be that the average Davis difference of about three adjusted raw-score units between experimentals and controls is simply not sufficient to generate measurable changes in writing. These changes, to the extent that the correlations between reading and writing fall below 1.00, could well be less in the first place--and of course pose far more severe measurement problems than are involved in assessing changes in reading competence. In addition, the "received wisdom" about writing changes is that they are notoriously slow. O'Hare's impressive gains (Sentence-Combining, NCTE, 1973) took place after a full year of instruction. Our own speculation is that since by comparison with reading, writing is a considerably more active and less automatic process, in the same way that talking is more active

than listening, there must be some urgency to do it. Nothing in the present experiment "encouraged" students to write more than perfunctorily. Earlier pilot results occurred in two kinds of situations: (1) clinical cases in which individuals (many of them university students or superior adults) with a strong incentive to write prior to S.O.S. appeared to experience considerable added competence shortly afterwards, (2) classes in which writing was a main issue, with substantial instruction in it, or very frequent occasions for doing it, and in which those who had had S.O.S. seemed to benefit more by the instruction or practice. Further research, therefore, ought to take the form of S.O.S.-plus-writing-instruction-and-practice versus non-S.O.S.-plus-comparable-writing-instruction-and-practice; and "writing instruction" should include lending status and importance to the act of writing. While some writing was obviously called for in the English classes of this experiment, there was no specific indication that intensive instruction or practice in it occurred between pretest and second posttest.

2. Although subjective measurements of quality were applied to only a small subsample (72) of the 194 students in the present writing experiment, other measures of writing were analyzed for the full sample, and it appears likely that the judgment of quality (and the Golub syntactic density measure, applied to about 50% of the sample) would hold for the whole sample. In the present experiment, a connection between S.O.S. and writing improvement has not been demonstrated.

3. The decline in "fluency" in both control and experimental students in the second posttest suggests that June is not a good time for the testing of free writing. In one experimental class, most students were observed to have written for only the first ten to twelve minutes of the thirty minutes that were provided.

R. G. Martin
J. F. Belanger

S.O.S. EXPERIMENT, NOVEMBER 1975 - JUNE 1976

DATA SHEET

The following summarizes means on Davis Series 2 Reading Tests with pretest immediately prior to administration of S.O.S. Reading Technique and posttests at intervals of 3 months and six months later. Forms used were D for pretest, B for first posttest, C for second posttest. As a rough way of equating the three forms, based on Davis's reported norms, taking Form D as a base, 2 raw score units can be added to scores on B and 1 raw score unit to scores on C. Scores reported here are actual raw scores except as noted. (All scoring by Davis R-$\frac{1}{4}$W formula.)

Experimental Students

	N	Nov,Dec (D)	Feb,Mar (B)	Gain	May,Jun (C)	Gain
Grade 9 (1) boys	(12)	20.750	21.083		24.083	
(1) girls	(10)	32.300	30.600		35.500	
Grade 9 (2) boys	(16)	40.062	40.500		39.000	
(2) girls	(14)	44.143	42.143		47.929	
Grade10 (1) boys	(10)	47.100	58.100		52.100	
(2) girls	(15)	52.400	55.600		54.000	
Grade10 (2) boys	(17)	46.765	50.059		49.412	
(2) girls	(6)	48.167	54.333		51.167	
Grade 9	(52)	35.211	34.558		37.289	
Grade 10	(48)	48.771	54.000		51.625	
Boys	(55)	39.200	42.418		41.345	
Girls	(45)	44.800	45.689		47.622	
Total (Exp.)	(100)	41.720	43.890	+2.170	44.170	+2.450

Control Students

	N	Nov,Dec (D)	Feb,Mar (B)	Gain	May,Jun (C)	Gain
Grade 9 (1) boys	(14)	41.143	41.286		41.357	
(1) girls	(12)	35.667	33.417		35.583	
Grade 9 (2) boys	(11)	46.455	44.000		48.273	
(2) girls	(18)	44.222	41.389		42.389	
Grade10 (1) boys	(14)	50.000	52.357		47.929	
(1) girls	(12)	53.750	53.583		50.667	
Grade10 (2) boys	(10)	50.300	50.800		49.700	
(2) girls	(14)	58.643	57.929		60.143	
Grade 9	(55)	42.018	40.146		41.818	
Grade 10	(50)	53.380	53.900		52.360	
Boys	(49)	46.734	47.000		46.490	
Girls	(56)	48.036	46.428		47.143	
Total (Controls)	(105)	47.429	46.695	-0.734	46.838	-0.591

Summary

	N	Nov,Dec (D)	Feb,Mar (B)	Gain	May,Jun (C)	Gain
Experimentals	(100)	41.720	43.890	+2.170	44.170	+2.450
Controls	(105)	47.429	46.695	-0.734	46.838	-0.591
Difference in favor of experimentals:				2.904		3.041
Same, with adjustment for estimated relative test form difficulty:						
Experimentals	(100)	41.720	45.890	+4.170	45.170	+3.450
Controls	(105)	47.429	48.695	+1.266	47.838	+0.409
Difference in favor of experimentals:				2.904		3.041

Significance of experimental-control difference, pretest to 1st posttest: .00232
" " " " " " 2nd " .00121

(Data analyzed by analysis of covariance, ANCOVA-35, Computer Centre, Faculty of Education, University of Alberta)

S.O.S. EXPERIMENT, NOVEMBER 1975 - JUNE 1976

DATA SHEET (PAGE 2)

The following summarizes means on writing samples collected at approximately the same time as Davis Reading tests were administered, as pretest, first posttest (or "midtest"), and second posttest (or simply "posttest"), in connection with the experimental administration of S.O.S. Reading Technique to ninth-grade and tenth-grade students. Topics ("The Ideal Home or Family", "The Ideal Country or Nation", "The Ideal School") were systematically rotated at the three testing periods within each class, so that each student wrote on all three, and equal numbers wrote on each topic at each administration. Each student had approximately 30 minutes in which to write each paper. Topic difficulty was estimated from rater evaluations of papers, and scores below reflect a correction for relative difficulty of the topics.

	Pretest	Midtest	Posttest
SUBJECTIVE EVALUATIONS BY RATERS (Subsample, 9 per class, N=72)			
Experimentals (N=36)	5.768	5.748	5.367
Controls (N=36)	5.787	6.098	5.728
MODIFIED GOLUB SYNTACTIC DENSITY (Subsample, 12 per class, N=96)			
Experimentals (N=48))	3.571	3.751	3.702
Controls (N=48)	3.741	3.931	3.689
WORDS PER T-UNIT (Entire sample, N=194)			
Experimentals (N=94)	14.744	14.687	14.704
Controls (N=100)	14.678	14.968	14.863
FLUENCY (WORDS PER PAPER) (Entire sample, N=194)			
Experimentals (N=94)	261.673	279.650	233.407
Controls (N=100)	270.520	261.401	243.202

Differences between controls and experimentals, analyzed by analysis of covariance, did not approach the .05 level. Further analysis of this phase (writing) of the S.O.S. experiment will be reported in the forthcoming thesis by J. Belanger.

APPENDIX B

Questions and Hypotheses

This study hypothesized that groups of experimental students who had shown significantly greater gains on reading measures than control groups had shown would, as a consequence, show significantly greater gains on writing measures than the control groups would show. To test this hypothesis, nine questions were asked to examine three grade groupings (the grade-nine students separately, the grade-ten students separately, and the grade-nine and -ten students combined) over three time periods (the first three months, the second three months, and the full six months of the experiment). The compositions of the experimental and control groups were compared at each grade level on four writing measures (quality, syntactic density, T-unit length, and fluency) for each of the three test periods. In addition, interactions between the treatment and the teacher who taught one of the four pairs of experimental and control classes involved in the study and the treatment and the sex of the subject were examined.

Since only three of the thirty-six null hypotheses tested in the study were rejected at the .05 level of confidence, a number of such rejections exceeding only slightly the number to be expected by chance alone, only a summary of these results was presented in Chapter V of this report. For the reader who wishes to examine the findings in detail, the complete statistical data for each of the thirty-six null hypotheses along with the interactions for each variable are presented in this appendix. For ease of

comparison, the tables presenting the raw scores on the reading measure[1] and the writing measures are also included here.

A. Examination of the Questions and Hypotheses

The nine questions asked in the study were presented in Chapter I. For convenience of the reader they are repeated here. The .05 level of confidence was considered statistically significant on all tests.

1. Grades Nine and Ten: The Pre- and Midtests. The first questions asked in the study compared the compositions of the grade-nine and grade-ten experimental and control students over the first three months of the study. The question asked was

> Do treatment classes of ninth- and tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown also show significantly greater growth in the quality and complexity of compositions written three months

---

[1]A summary of the reading analysis was presented in Chapter IV. As was noted in Chapter IV, the reading changes were obtained as part of a concurrent experiment. The quality and syntactic density analyses, as reported in Chapter III, used 37- and 50-percent random samples of the experimental population respectively while the T-unit and fluency analyses used the total experimental population. Consequently, the scores on the reading measure for the two samples differ from the scores for the experimental population, although, as noted in Chapter IV, these differences were, for the most part, statistically non-significant. Therefore, separate reading tables are presented for the quality, syntactic density, and T-unit analyses. The reading scores are the same for the T-unit and fluency analyses and, consequently, the reading tables are not repeated for the nine null hypotheses examining the fluency scores.

> following the administration of the treatment when the compositions of the two groups are judged by experienced raters and when they are examined for syntactic density, average T-unit length, and writing fluency?

For purposes of statistical analysis this question was stated in terms of null hypotheses examining the quality, syntactic density, average T-unit length, and fluency of the compositions.

> Null Hypothesis 1a: Compositions written by treatment classes of ninth- and tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be judged significantly different in quality from those written by non-treatment classes when the compositions are written three months following the administration of the treatment.

TABLE 41

Pre-mid comparison on the quality measure for grades nine and ten combined.

Source	df*	F-ratio	probability	decision
Treatment	1	.0014	.97	N.S.**
Treatment - Teacher	3	.831	.48	N.S.
Treatment - Sex	1	.228	.64	N.S.

*degrees of freedom

**not statistically significant at the .05 level of confidence

As Table 41 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the quality of the compositions written by the ninth- and tenth-grade experimental and control students was

not significantly different three months following the administration of the treatment when initial differences between the two groups were adjusted by the procedure of analysis of covariance. The very high probability of .97 indicates that measurable differences between the two groups were indeed very slight.

Table 41 also shows that there was no significant interaction between the treatment and the teacher who taught one of the four pairs of experimental and control classes involved in the study or between the treatment and the sex of the subject. The results of the quality analysis, then, were not a factor of any pedagogical influence specific to any one of the four teachers in the study, nor were they significantly influenced by the sex of the subject.[1]

As Table 42 shows, the changes in writing quality were not in the hypothesized direction. Over the first three months of the experiment slight changes in writing quality favored the control group. Table 43 presents the pre- and midtest reading scores of the samples used in the quality analysis.

> Null Hypothesis 1b: Compositions written by treatment classes of ninth- and tenth-grade students who have shown significantly greater gains on reading measures than non-

---

[1]No significant interaction between the treatment and the teacher was found in this study. Consequently, although the statistical results are given in the tables, these interactions will not be discussed in the text of the appendix. Twenty-percent of the interactions between the treatment and the sex of the subject were statistically significant. These were all attributable to the scores of the control girls and will be noted when relevant.

TABLE 42

Pre-mid comparison: average raw scores on the writing quality measure for grades nine and ten combined.

	n	Pretest Raw Scores $\bar{x}$	σ	Midtest Raw Scores $\bar{x}$	σ	Midtest Minus Pretest
Experimental						
Boys	21	5.81	1.29	5.56	1.10	-0.25
Girls	15	5.71	1.17	6.02	1.23	+0.31
Combined	36	5.77	1.24	5.75	1.18	-0.02
Control						
Boys	12	6.10	1.68	5.93	1.75	-0.17
Girls	24	5.63	1.63	6.19	1.95	+0.56
Combined	36	5.79	1.66	6.10	1.89	+0.31

TABLE 43

Pre-mid comparison: average raw scores on the reading measure for the grade-nine and -ten samples used in the quality analysis.

	n	Pretest Raw Scores $\bar{x}$	σ	Midtest Raw Scores $\bar{x}$	σ	Midtest Minus Pretest
Experimental						
Boys	21	41.52	15.21	46.52	16.40	+5.00
Girls	15	41.00	13.47	40.27	13.50	-0.73
Combined	36	41.31	14.51	43.92	15.57	+2.61
Control						
Boys	12	47.25	13.14	46.58	13.27	-0.67
Girls	24	47.67	15.82	45.54	19.01	-2.13
Combined	36	47.53	14.98	45.89	17.32	-1.64

treatment classes have shown will not be significantly different in syntactic density from those written by non-treatment classes when the compositions are written three months following the administration of the treatment.

As Table 44 shows, at the .05 level of confidence the

TABLE 44

Pre-mid comparison on the syntactic density measure for grades nine and ten combined.

Source	df*	F-ratio	probability	decision
Treatment	1	.483	.49	N.S.**
Treatment - Teacher	3	.693	.56	N.S.
Treatment - Sex	1	.498	.48	N.S.

*degrees of freedom

**not statistically significant at the .05 level of confidence

null hypothesis was not rejected and it is therefore concluded that the syntactic density of the compositions written by the ninth- and tenth-grade experimental and control students was not significantly different three months following the administration of the treatment when initial differences between the two groups were adjusted by the procedure of analysis of covariance. The high probability of .49 suggests that measurable differences between the two groups were small.

As Table 45 shows, the changes in syntactic density were not in the hypothesized direction. Over the first three months of the experiment, the experimental and control groups showed almost identical changes in syntactic density scores. Table 46 presents the pre- and midtest reading scores of the samples used in the syntactic density analysis.

Null Hypothesis 1c: Compositions written by treatment

TABLE 45

Pre-mid comparisons: average raw scores on the syntactic density measure for grades nine and ten combined.

	n	Pretest Raw Scores $\bar{x}$	σ	Midtest Raw Scores $\bar{x}$	σ	Midtest Minus Pretest
Experimental						
Boys	27	3.62	0.83	3.83	0.80	+0.21
Girls	21	3.51	0.75	3.65	0.88	+0.14
Combined	48	3.57	0.80	3.75	0.84	+0.18
Control						
Boys	18	3.94	1.18	4.06	1.04	+0.12
Girls	30	3.62	0.90	3.85	0.98	+0.23
Combined	48	3.74	1.03	3.93	1.01	+0.19

TABLE 46

Pre-mid comparison: average raw scores on the reading measure for the grade-nine and -ten samples used in the syntactic density analysis.

	n	Pretest Raw Scores $\bar{x}$	σ	Midtest Raw Scores $\bar{x}$	σ	Midtest Minus Pretest
Experimental						
Boys	27	40.11	16.34	43.82	17.42	+3.71
Girls	21	44.48	14.00	43.14	15.19	-1.34
Combined	48	42.02	15.51	43.52	16.48	+1.50
Control						
Boys	18	47.39	13.65	47.28	13.80	-0.11
Girls	30	49.70	15.57	47.13	18.22	-2.57
Combined	48	48.83	14.92	47.19	16.70	-1.64

classes of ninth- and tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in average T-unit length from those written by non-treatment classes when the compositions are written three months following the administration of the treatment.

TABLE 47

Pre-mid comparison on the T-unit measure for grades nine and ten combined.

Source	df	F-ratio	probability	decision
Treatment	1	0.998	.32	N.S.
Treatment - Teacher	3	2.585	.06	N.S.
Treatment - Sex	1	3.997	.05	Sig

As Table 47 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the average T-unit length of the compositions written by the ninth- and tenth-grade experimental and control students was not significantly different three months following the administration of the treatment when initial differences between the two groups were adjusted by the procedure of analysis of covariance. The probability of .32 indicates that measurable differences between the two groups were small.

Table 47 also shows that at the .05 level of confidence there was a significant interaction between the treatment and the sex of the subject.

Table 48 shows that the changes in average T-unit length were not in the hypothesized direction. The control group showed gains on this measure over the first three months of the experiment while the experimental group showed slight losses. Table 48 also shows that the control girls,

TABLE 48

Pre-mid comparison: average raw scores on the T-unit measure for grades nine and ten combined.

	n	Pretest Raw Scores $\bar{x}$	σ	Midtest Raw Scores $\bar{x}$	σ	Midtest Minus Pretest
Experimental						
Boys	53	14.44	3.06	14.65	2.89	+0.21
Girls	41	15.13	3.06	14.73	3.43	-0.40
Combined	94	14.74	3.08	14.69	3.14	-0.05
Control						
Boys	47	15.00	3.12	15.07	3.03	+0.07
Girls	53	14.39	2.80	14.88	3.16	+0.49
Combined	100	14.68	2.97	14.97	3.10	+0.29

TABLE 49

Pre-mid comparison: average raw scores on the reading measure for the grade-nine and -ten samples used in the T-unit and fluency analyses.

	n	Pre-test Raw Scores $\bar{x}$	σ	Midtest Raw Scores $\bar{x}$	σ	Midtest Minus Pretest
Experimental						
Boys	53	39.60	16.36	42.70	18.18	+3.10
Girls	41	45.44	16.36	45.90	17.13	+0.46
Combined	94	42.15	16.62	44.10	17.80	+1.95
Control						
Boys	47	47.11	14.05	46.89	13.97	-0.22
Girls	53	47.74	16.46	46.23	18.02	-1.51
Combined	100	47.44	15.37	46.54	16.25	-0.90

the subgroup recording the shortest average T-unit lengths on the pretest, showed the largest gains on the midtest. Combined with losses by the experimental girls, the subgroup recording the longest average T-unit lengths on the pretest

and showing the largest losses on the midtest, this produced the significant interaction between the treatment and the sex of the subject noted above. Table 49 presents the pre- and midtest reading scores of the experimental population used in both the T-unit analysis and fluency analysis (below).

> Null Hypothesis 1d: Compositions written by treatment classes of ninth- and tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in fluency from those written by non-treatment classes when the compositions are written three months following the administration of the treatment.

TABLE 50

Pre-mid comparison on the fluency measure for grades nine and ten combined.

Source	df	F-ratio	probability	decision
Treatment	1	1.714	.19	N.S.
Treatment - Teacher	3	0.469	.71	N.S.
Treatment - Sex	1	0.848	.36	N.S.

As Table 50 indicates, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the fluency of the compositions written by the ninth- and tenth-grade experimental and control students was not significantly different three months following the administration of the treatment when initial differences between the two groups were adjusted by the procedure of

analysis of covariance. The probability of .19 indicates that measurable differences between the two groups were small.

TABLE 51

Pre-mid comparison: average raw scores on the fluency measure for grades nine and ten combined.

	n	Pretest Raw Scores $\bar{x}$	σ	Midtest Raw Scores $\bar{x}$	σ	Midtest Minus Pretest
Experimental						
Boys	53	252.3	100.5	271.7	78.3	+19.4
Girls	41	273.8	100.1	289.9	98.5	+16.1
Combined	94	261.7	100.9	279.6	88.1	+17.9
Control						
Boys	47	280.8	95.2	269.7	91.0	-11.1
Girls	53	261.4	98.2	254.0	100.5	-7.4
Combined	100	270.5	97.3	261.4	96.5	-9.1

As Table 51 shows, the changes on the fluency measure were in the hypothesized direction. Over the first three months of the experiment the experimental group showed gains on this measure while the control group showed losses. These differences, however, did not approach statistical significance.

2. Grade Nine: The Pre- and Midtests. The second question asked in the study compared the compositions of the grade-nine experimental and control students over the first three months of the study.

> Do treatment classes of ninth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown also show significantly greater growth in the quality and complexity of compositions written three months

> following the administration of the treatment when the compositions are judged by experienced raters and when they are examined for syntactic density, T-unit length, and writing fluency?

For purposes of statistical analysis this question was stated in terms of null hypotheses examining the quality, syntactic density, average T-unit length, and fluency of the compositions.

> Null Hypothesis 2a: Compositions written by treatment classes of ninth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be judged significantly different in quality from those written by non-treatment classes when the compositions are written three months following the administration of the treatment.

TABLE 52

Pre-mid comparison on the writing quality measure for grade-nine subjects.

Source	df	F-ratio	probability	decision
Treatment	1	.337	.57	N.S.
Treatment - Teacher	1	.151	.70	N.S.
Treatment - Sex	1	.032	.86	N.S.

As Table 52 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the quality of the compositions written by the ninth-grade experimental and control students was not significantly different three months following the administration of the treatment when initial differences between the two groups

were adjusted by the procedure of analysis of covariance. The high probability of .57 suggests that measurable differences between the two groups were small.

TABLE 53

Pre-mid comparison: average raw scores on the quality measure for grade-nine subjects.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Midtest Minus Pretest
Experimental						
Boys	8	6.16	1.32	5.95	0.90	-.21
Girls	10	5.22	0.72	5.74	1.23	+.52
Combined	18	5.64	1.13	5.83	1.10	+.19
Control						
Boys	6	5.64	2.15	5.77	1.04	+.13
Girls	12	5.30	1.72	6.04	1.25	+.74
Combined	18	5.41	1.88	5.95	1.19	+.54

TABLE 54

Pre-mid comparison: average raw scores on the reading measure for the grade-nine sample used in the quality analysis.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Midtest Minus Pretest
Experimental						
Boys	8	33.38	11.77	35.25	11.26	+1.87
Girls	10	39.90	12.72	38.50	13.43	-1.40
Combined	18	37.00	12.73	37.06	12.62	+0.06
Control						
Boys	6	46.67	12.26	44.33	12.71	-2.34
Girls	12	41.42	13.43	37.67	15.53	-3.75
Combined	18	43.17	13.28	39.89	14.98	-3.28

As Table 53 shows, the changes in writing quality

were not in the hypothesized direction. Over the first three months of the experiment the control group showed greater gains on this measure than the experimental group did. Table 54 presents the pre- and midtest reading scores of the grade-nine samples used in the quality analysis.

> Null hypothesis 2b: Compositions written by treatment classes of ninth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in syntactic density from those written by the non-treatment classes when the compositions are written three months following the administration of the treatment.

TABLE 55

Pre-mid comparison on the syntactic density measure for grade-nine subjects.

Source	df	F-ratio	probability	decision
Treatment	1	0.585	.45	N.S.
Treatment - Teacher	1	1.020	.32	N.S.
Treatment - Sex	1	0.106	.75	N.S.

As Table 55 indicates, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the syntactic density of the compositions written by the ninth-grade experimental and control students was not significantly different three months following the administration of the treatment when initial differences between the two groups were adjusted by the procedure of analysis of covariance. The high probability of .45 suggests

that measurable differences between the two groups were slight.

TABLE 56

Pre-mid comparison: average raw scores on the syntactic density measure for grade-nine subjects.

	n	Pretest Raw Scores $\bar{x}$	σ	Midtest Raw Scores $\bar{x}$	σ	Midtest Minus Pretest
Experimental						
Boys	11	3.34	.61	3.73	0.67	+.39
Girls	13	3.64	.69	3.71	0.83	+.07
Combined	24	3.51	.67	3.72	0.76	+.21
Control						
Boys	9	3.45	.45	3.69	1.17	+.24
Girls	15	3.86	.83	3.65	0.86	-.21
Combined	24	3.71	.74	3.67	0.99	-.04

TABLE 57

Pre-mid comparison: average raw scores on the reading measure for the grade-nine sample used in the syntactic density analysis.

	n	Pretest Raw Scores $\bar{x}$	σ	Midtest Raw Scores $\bar{x}$	σ	Midtest Minus Pretest
Experimental						
Boys	11	28.82	12.90	30.73	13.44	+1.91
Girls	13	42.69	14.37	39.31	15.34	-3.38
Combined	24	36.33	15.36	35.38	15.12	-0.95
Control						
Boys	9	44.56	12.93	43.44	13.53	-1.12
Girls	15	44.67	14.81	40.27	16.39	-4.40
Combined	24	44.63	14.13	41.46	15.45	-3.17

Table 56 indicates that the changes in syntactic density were in the hypothesized direction. Over the first

three months of the experiment, the grade-nine experimental group showed gains on this measure while the grade-nine control group showed slight losses. However, as noted above, these differences were slight. Table 57 presents the pre- and midtest reading scores of the grade-nine samples used in the syntactic density analysis.

> Null Hypothesis 2c: Compositions written by treatment classes of ninth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in average T-unit length from those written by non-treatment classes when the compositions are written three months following the administration of the treatment.

TABLE 58

Pre-mid comparison on the T-unit measure for the grade-nine subjects.

Source	df	F-ratio	probability	decision
Treatment	1	.635	.43	N.S.
Treatment - Teacher	1	.0001	.99	N.S.
Treatment - Sex	1	.079	.78	N.S.

As Table 58 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the average T-unit length of the compositions written by the ninth-grade experimental and control students was not significantly different three months following the administration of the treatment when initial differences between the two groups were adjusted by the procedure of

analysis of covariance. The probability of .43 indicates that measurable differences between the two groups were small.

TABLE 59

Pre-mid comparison: average raw scores on the T-unit measure for grade-nine subjects.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Midtest Minus Pretest
Experimental						
Boys	28	13.73	2.92	14.38	2.77	+0.65
Girls	22	14.29	2.70	14.09	3.47	-0.20
Combined	50	13.97	2.84	14.25	3.10	+0.28
Control						
Boys	24	13.93	2.23	14.21	2.78	+0.28
Girls	29	15.00	2.95	14.16	2.58	-0.84
Combined	53	14.51	2.70	14.18	2.67	-0.33

TABLE 60

Pre-mid comparison: average raw scores on the reading measure for the grade-nine samples used in the T-unit and fluency analyses.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Midtest Minus Pretest
Experimental						
Boys	28	31.79	15.47	32.19	15.60	+0.40
Girls	22	40.41	13.28	37.82	14.70	-2.59
Combined	50	35.58	15.16	34.66	15.47	-0.92
Control						
Boys	24	44.25	14.47	42.75	14.37	-1.50
Girls	29	41.45	14.27	38.79	14.42	-2.66
Combined	53	42.72	14.43	40.59	14.53	-2.13

Table 59 shows that differences in average T-unit length were in the hypothesized direction. Over the first

three months of the experiment, the experimental group showed increases in average T-unit length while the control group showed decreases. Again, these differences were small. Table 60 presents the pre- and midtest reading scores of the grade-nine sample used in both the T-unit analysis and the fluency analysis (below).

> Null Hypothesis 2d: Compositions written by treatment classes of ninth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in fluency from those written by non-treatment classes when the compositions are written three months following the administration of the treatment.

TABLE 61

Pre-mid comparison on the fluency measure for grade-nine subjects.

Source	df	F-ratio	probability	decision
Treatment	1	2.494	.12	N.S.
Treatment - Teacher	1	1.825	.18	N.S.
Treatment - Sex	1	0.008	.93	N.S.

As Table 61 indicates, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the fluency of the compositions written by the ninth-grade experimental and control students was not significantly different three months following the administration of the treatment when initial differences between the two groups were adjusted by the procedure of

analysis of covariance. The probability of .12 suggests that measurable differences between the two groups were not substantial.

TABLE 62

Pre-mid comparison: average raw scores on the fluency measure for grade-nine subjects.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Midtest Minus Pretest
Experimental						
Boys	28	241.3	91.4	258.4	80.9	+17.1
Girls	22	250.7	102.3	265.3	106.5	+14.6
Combined	50	245.4	96.5	261.4	93.1	+16.0
Control						
Boys	24	256.7	71.6	250.0	110.2	-6.7
Girls	29	275.4	86.0	262.2	77.7	-13.2
Combined	53	266.9	80.3	256.6	94.0	-10.3

Table 62 shows that the changes in fluency for the grade-nine sample, although statistically non-significant, were in the hypothesized direction. Over the first three months of the experiment, the experimental students showed gains on the fluency measure while the control students showed losses.

3. Grade Ten: The Pre- and Midtests. The third question asked in the study compared the writing changes of the grade-ten experimental and control students over the first three months of the study. The question asked was

> Do treatment classes of tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown also show

significantly greater growth in the quality and complexity of compositions written three months following the administration of the treatment when the compositions of the two groups are judged by experienced raters and when they are examined for syntactic density, average T-unit length, and writing fluency?

For purposes of statistical analysis this question was stated in terms of null hypotheses examining the quality, syntactic density, average T-unit length, and fluency of the compositions.

Null Hypothesis 3a: Compositions written by treatment classes of tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be judged significantly different in quality from those written by non-treatment classes when the compositions are written three months following the administration of the treatment.

TABLE 63

Pre-mid comparison on the quality measure for grade-ten subjects.

Source	df	F-ratio	probability	decision
Treatment	1	0.284	.60	N.S.
Treatment - Teacher	1	1.081	.31	N.S.
Treatment - Sex	1	1.239	.28	N.S.

As Table 63 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the quality of the compositions written by the tenth-grade experimental and control students was not significantly different three months following the administration of the treatment when initial differences

between the two groups were adjusted by the procedure of analysis of covariance. The high probability of .60 indicates that measurable differences between the two groups were slight.

TABLE 64

Pre-mid comparison: average raw scores on the quality measure for grade-ten subjects.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Midtest Minus Pretest
Experimental						
Boys	13	5.60	1.23	5.32	1.13	-0.28
Girls	5	6.68	1.30	6.57	1.02	-0.11
Combined	18	5.90	1.34	5.66	1.24	-0.24
Control						
Boys	6	6.57	0.76	6.08	2.24	-0.49
Girls	12	5.96	1.47	6.33	2.44	+0.37
Combined	18	6.16	1.31	6.25	2.38	+0.09

TABLE 65

Pre-mid comparison: average raw scores on the reading measure for the grade-ten sample used in the quality analysis.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Midtest Minus Pretest
Experimental						
Boys	13	46.54	14.91	53.46	15.18	+6.92
Girls	5	43.20	14.59	43.80	12.92	+0.60
Combined	18	45.61	13.90	50.78	15.22	+5.17
Control						
Boys	6	47.83	13.95	48.83	13.43	+1.00
Girls	12	53.92	15.56	53.42	18.91	-0.50
Combined	18	51.89	15.32	51.89	17.42	0

Table 64 shows that the changes in writing quality

were not in the hypothesized direction. Over the first three months of the experiment the grade-ten control groups registered slight gains in writing quality while the grade-ten experimental groups registered slight losses. Table 65 presents the pre- and midtest reading scores of the grade-ten sample used in the quality analysis.

> Null Hypothesis 3b: Compositions written by treatment classes of tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in syntactic density from those written by the non-treatment classes when the compositions are written three months following the administration of the treatment.

TABLE 66

Pre-mid comparison on the syntactic density measure for grade-ten subjects.

Source	df	F-ratio	probability	decision
Treatment	1	0.038	.85	N.S.
Treatment - Teacher	1	1.036	.32	N.S.
Treatment - Sex	1	0.912	.35	N.S.

As Table 66 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the syntactic density of the compositions written by the tenth-grade experimental and control students was not significantly different three months following the administration of the treatment when initial differences between the two groups were adjusted by the procedure of

analysis of covariance. The very high probability of .85 suggests that measurable differences between the two groups were indeed slight.

TABLE 67

Pre-mid comparison: average raw scores on the syntactic density measure for grade-ten subjects.

	n	Pretest Raw Scores $\bar{x}$	σ	Midtest Raw Scores $\bar{x}$	σ	Midtest Minus Pretest
Experimental						
Boys	16	3.81	0.90	3.90	0.87	+0.09
Girls	8	3.29	0.79	3.57	0.96	+0.28
Combined	24	3.64	0.90	3.79	0.92	+0.15
Control						
Boys	9	4.44	1.45	4.43	0.73	-0.01
Girls	15	3.38	0.91	4.05	1.04	+0.67
Combined	24	3.78	1.25	4.19	0.95	+0.41

TABLE 68

Pre-mid comparison: average raw scores on the reading measure for the grade-ten sample used in the syntactic density analysis.

	n	Pretest Raw Scores $\bar{x}$	σ	Midtest Raw Scores $\bar{x}$	σ	Midtest Minus Pretest
Experimental						
Boys	16	47.88	13.72	52.81	13.75	+4.93
Girls	8	47.38	12.85	49.38	12.66	+2.00
Combined	24	47.71	13.44	51.67	13.50	+3.96
Control						
Boys	9	50.22	13.76	51.11	12.97	-0.89
Girls	15	54.73	14.66	54.00	17.36	-0.73
Combined	24	53.04	14.50	52.92	15.92	-0.12

Table 67 shows that changes on the syntactic density

measure were not in the hypothesized direction. Over the first three months of the experiment the control group registered slightly greater gains on the syntactic density measure than the experimental group did. Table 68 presents the pre- and midtest reading scores of the grade-ten sample used in the syntactic density analysis.

> Null Hypothesis 3c: Compositions written by treatment classes of tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in average T-unit length from those written by non-treatment classes when the compositions are written three months following the administration of the treatment.

TABLE 69

Pre-mid comparison on the T-unit measure for grade-ten subjects.

Source	df	F-ratio	probability	decision
Treatment	1	4.236	.04	Sig
Treatment - Teacher	1	3.065	.08	N.S.
Treatment - Sex	1	7.334	.008	Sig

As Table 69 shows, at the .05 level of confidence the null hypothesis was rejected and it is therefore concluded that the average T-unit length of the compositions written by the tenth-grade experimental and control students was significantly different three months following the administration of the treatment when initial differences between the two groups were adjusted by the procedure of

analysis of covariance. Table 69 also shows that there was a significant interaction between the treatment and the sex of the subject.

TABLE 70

Pre-mid comparison: average raw scores on the T-unit measure for grade-ten subjects.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Midtest Minus Pretest
Experimental						
Boys	25	15.24	3.02	14.96	2.99	-0.28
Girls	19	16.11	3.16	15.47	3.24	-0.64
Combined	44	15.62	3.11	15.18	3.11	-0.44
Control						
Boys	23	16.13	3.51	15.97	3.02	-0.16
Girls	24	13.66	2.42	15.75	3.54	+2.09
Combined	47	14.87	3.24	15.86	3.30	+0.99

TABLE 71

Pre-mid comparison: average raw scores on the reading measure for grade-ten samples used in the T-unit and fluency analyses.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Midtest Minus Pretest
Experimental						
Boys	25	48.36	12.44	54.48	12.85	+6.12
Girls	19	51.26	17.61	55.26	14.83	+4.00
Combined	44	49.61	14.96	54.82	13.74	+5.21
Control						
Boys	23	50.09	12.93	51.22	12.12	+1.13
Girls	24	55.33	15.70	55.21	17.84	-0.12
Combined	47	52.77	14.65	53.26	15.44	+0.49

As Table 70 shows, the changes in average T-unit

length were not in the hypothesized direction. The control students showed gains on this measure while the experimental students showed losses. Inspection of Table 70 reveals that the control girls accounted for almost all of these differences. The control girls, who scored over one standard deviation below the group mean on the pretest, scored near the mean on the midtest. This appears, then, to be a result of regression towards the mean rather than the result of real growth. It should be noted, too, that this factor influenced the significant interaction between treatment and sex on the T-unit measure when the grade-nine and -ten groups were combined (cf. Null Hypothesis 1c).

Table 71 presents the pre- and midtest reading scores of the grade-ten samples used in the T-unit length and fluency analyses.

> Null Hypothesis 3d: Compositions written by treatment classes of tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in fluency from those written by non-treatment classes when the compositions are written three months following the administration of the treatment.

As Table 72 indicates, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the fluency of the compositions written by the tenth-grade experimental and control students was not significantly different three months following the administration of the treatment when initial differences between the two groups were adjusted by the procedure of analysis of covariance. The probability of .39 indicates

TABLE 72

Pre-mid comparison on the fluency measure for grade-ten subjects.

Source	df	F-ratio	probability	decision
Treatment	1	.756	.39	N.S.
Treatment - Teacher	1	.0002	.99	N.S.
Treatment - Sex	1	.902	.35	N.S.

that measurable differences between the experimental and control groups were small.

TABLE 73

Pre-mid comparison: average raw scores on the fluency measure for grade-ten subjects.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Midtest Minus Pretest
Experimental						
Boys	25	264.7	108.4	286.7	72.4	+22.0
Girls	19	300.4	90.5	318.4	79.3	+18.0
Combined	44	280.1	102.6	300.4	77.0	+20.3
Control						
Boys	23	305.9	109.2	290.4	58.4	-15.5
Girls	24	244.5	108.8	244.1	121.9	-.4
Combined	47	274.6	113.2	266.8	98.9	-7.8

Table 73 shows that the changes in fluency for the grade-ten sample, although statistically non-significant, were in the hypothesized direction. The experimental students showed gains on the fluency measure while the

control students showed losses.

4. Grades Nine and Ten Combined: The Mid- and Posttests. The fourth question asked in the study compared the writing changes of the grade-nine and grade-ten experimental and control students over the second three months of the study. The question asked was

> Do treatment classes of ninth- and tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown also show significantly greater growth in the quality and complexity of compositions written three and six months following the administration of the treatment when the compositions of the two groups are judged by experienced raters and when they are examined for syntactic density, average T-unit length, and writing fluency?

For purposes of statistical analysis this question was stated in terms of null hypotheses examining the quality, syntactic density, T-unit length, and fluency of the compositions.

> Null Hypothesis 4a: Compositions written by treatment classes of ninth- and tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be judged significantly different in quality from those written by non-treatment classes when the compositions written three months following the administration of the treatment are compared with those written six months following the administration of the treatment.

As Table 74 shows, at the .05 level of confidence, the null hypothesis was not rejected and it is therefore concluded that the quality of the compositions written by the ninth- and tenth-grade experimental and control students was not significantly different when the compositions written three months following the administration of the treatment were compared with those written six months

TABLE 74

Mid-post comparison on the quality measure for grades nine and ten combined.

Source	df	F-ratio	probability	decision
Treatment	1	.002	.96	N.S.
Treatment - Teacher	3	.521	.67	N.S.
Treatment - Sex	1	.012	.91	N.S.

following the administration of the treatment and when the initial differences between the two groups were adjusted by the procedure of analysis of covariance. The very high probability of .96 indicates that measurable differences between the two groups were indeed very slight.

Table 75 shows that over the second three months of the experiment the changes in writing quality were not in the hypothesized direction. The experimental and control groups showed small but equal losses on the writing quality measure. Table 76 presents the mid- and posttest reading scores of the samples used in the quality analysis.

> Null Hypothesis 4B: Compositions written by treatment classes of ninth- and tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in syntactic density from those written by non-treatment classes when the compositions written three months following the administration of the treatment are compared with those written six months following the administration of the treatment.

As Table 77 shows, at the .05 level of confidence, the null hypothesis was not rejected and it is therefore

TABLE 75

Mid-post comparison: average raw scores on the quality measure for grades nine and ten combined.

	n	Midtest Raw Scores $\bar{x}$	σ	Posttest Raw Scores $\bar{x}$	σ	Posttest Minus Midtest
Experimental						
Boys	21	5.56	1.10	5.11	1.63	-.45
Girls	15	6.02	1.23	5.73	1.03	-.29
Combined	36	5.75	1.18	5.37	1.45	-.38
Control						
Boys	12	5.93	1.75	5.72	1.68	-.21
Girls	24	6.19	1.95	5.74	1.75	-.45
Combined	36	6.10	1.89	5.73	1.72	-.37

TABLE 76

Mid-post comparison: average raw scores on the reading measure for the grade-nine and -ten samples used in the quality analysis.

	n	Midtest Raw Scores $\bar{x}$	σ	Posttest Raw Scores $\bar{x}$	σ	Posttest Minus Midtest
Experimental						
Boys	21	46.52	16.40	41.71	17.12	-4.81
Girls	15	40.27	13.50	41.47	14.63	+1.20
Combined	36	43.92	15.57	41.61	16.13	-2.31
Control						
Boys	12	46.58	13.27	47.08	16.84	+0.50
Girls	24	45.54	19.01	44.38	18.40	-1.16
Combined	36	45.89	17.32	45.28	17.94	-0.61

concluded that the syntactic density of the compositions written by the ninth- and tenth-grade experimental and control students was not significantly different when the compositions written three months following the administration

TABLE 77

Mid-post comparison on the syntactic density measure for grades nine and ten combined.

Source	df	F-ratio	probability	decision
Treatment	1	1.185	.18	N.S.
Treatment - Teacher	3	0.871	.46	N.S.
Treatment - Sex	1	0.788	.38	N.S.

of the treatment were compared with those written six months following the administration of the treatment and when the initial differences between the two groups were adjusted by the procedure of analysis of covariance. The probability of .18 indicates that measurable differences between the two groups were small.

TABLE 78

Mid-post comparison: average raw scores on the syntactic density measure for grades nine and ten combined.

		Midtest Raw Scores		Posttest Raw Scores		Posttest Minus Midtest
	n	$\bar{x}$	σ	$\bar{x}$	σ	
Experimental						
Boys	27	3.83	0.80	3.88	.80	+.05
Girls	21	3.65	0.88	3.48	.70	-.17
Combined	48	3.75	0.84	3.70	.79	-.05
Control						
Boys	18	4.06	1.04	3.50	.93	-.56
Girls	30	3.85	0.98	3.80	.91	-.05
Combined	48	3.93	1.01	3.69	.93	-.24

Table 78 shows that the changes in average T-unit length were in the hypothesized direction. The experimental

TABLE 79

Mid-post comparison: average raw scores on the reading measure for the grade-nine and -ten samples used in the syntactic density analysis.

	n	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Midtest
Experimental						
Boys	27	43.82	17.42	39.89	17.82	-3.93
Girls	21	43.14	15.19	44.68	15.25	+1.54
Combined	48	43.52	16.48	41.98	16.91	-1.54
Control						
Boys	18	47.28	13.80	46.78	15.62	-0.50
Girls	30	47.13	18.22	46.40	18.33	-0.73
Combined	48	47.19	16.70	46.54	17.37	-0.65

group showed fewer losses on the syntactic density measure over the second three months of the experiment than the control students did. Table 79 presents the mid- and posttest reading scores of the samples used in the syntactic density analysis.

> Null Hypothesis 4c: Compositions written by treatment classes of ninth- and tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in average T-unit length from those written by non-treatment classes when the compositions written three months following the administration of the treatment are compared with those written six months following the administration of the treatment.

As Table 80 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the average T-unit length of the compositions written by the ninth- and tenth-grade experimental and control students was not significantly different when the

[illegible]

[illegible]

[illegible]

[illegible]

TABLE 80

Mid-post comparison on the T-unit measure for grades nine and ten combined.

Source	df	F-ratio	probability	decision
Treatment	1	0.757	.39	N.S.
Treatment - Teacher	3	2.100	.10	N.S.
Treatment - Sex	1	0.137	.71	N.S.

compositions written three months following the administration of the treatment were compared with those written six months following the administration of the treatment and when the initial differences between the two groups were adjusted by the procedure of analysis of covariance. The probability of .39 indicates that measurable differences between the two groups were small.

Table 81 shows that although the differences between the experimental and control groups were not significant, they were in the hypothesized direction. The experimental group showed slight gains on T-unit measure while the control group showed slight losses. Table 81 presents the mid- and posttest reading scores of the samples used in both the T-unit analysis and fluency analysis (below).

> <u>Null Hypothesis 4d</u>: Compositions written by treatment classes of ninth- and tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in fluency from those written by the non-treatment classes when the compositions written three months following the administration of the treatment are compared with those written six months

TABLE 81

Mid-post comparison: average raw scores on the T-unit measure for grades nine and ten combined.

	n	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Midtest
Experimental						
Boys	53	14.65	2.89	14.70	2.66	+.05
Girls	41	14.73	3.43	14.71	3.49	-.02
Combined	94	14.69	3.14	14.70	3.05	+.01
Control						
Boys	47	15.07	3.03	14.49	3.02	-.58
Girls	53	14.88	3.16	15.20	3.36	+.32
Combined	100	14.97	3.10	14.86	3.22	-.11

TABLE 82

Mid-post comparison: average raw scores on the reading measure for the grade-nine and -ten samples used in the T-unit and fluency analyses.

	n	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Midtest
Experimental						
Boys	53	42.70	18.18	41.74	17.30	-0.96
Girls	41	45.90	17.13	47.76	16.69	+1.86
Combined	94	44.10	17.80	44.36	17.29	+0.26
Control						
Boys	47	46.89	13.97	46.38	15.04	-0.51
Girls	53	46.23	18.02	47.06	17.75	+0.83
Combined	100	46.54	16.25	46.74	16.54	+0.20

following the administration of the treatment.

As Table 83 shows, at the .05 level of confidence the null hypothesis was rejected and it is therefore concluded that the fluency of the compositions written by

TABLE 83

Mid-post comparison on the fluency measure for grades nine and ten combined.

Source	df	F-ratio	probability	decision
Treatment	1	4.377	.04	Sig
Treatment - Teacher	3	0.763	.52	N.S.
Treatment - Sex	1	1.852	.18	N.S.

the ninth- and tenth-grade experimental and control students was significantly different when the compositions written three months following the administration of the treatment were compared with those written six months following the administration of the treatment and when initial differences between the two groups were adjusted by the procedure of analysis of covariance. These differences, however, favored the control groups. As was noted in Null Hypothesis 1d, the experimental group showed gains in fluency over the first three months of the experiment while the control group showed losses. These differences, however, were not significant. Over the second three months of the experiment, the experimental students lost over twice the gains they had shown over the first three months while the control students again showed modest losses. Over the full six months of the experiment, as will be noted in Null Hypothesis 7d, the experimental and control students showed equal losses.

Table 84 shows that the changes on the fluency

TABLE 84

Mid-post comparison: average raw scores on the fluency measure for grades nine and ten combined.

	n	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Midtest
Experimental						
Boys	53	271.7	78.3	225.2	92.4	-46.5
Girls	41	289.9	98.5	244.1	87.1	-45.8
Combined	94	279.6	88.1	233.4	90.6	-46.2
Control						
Boys	47	269.7	91.0	241.0	94.4	-28.7
Girls	53	254.0	100.5	245.2	93.8	-8.8
Combined	100	261.4	96.5	243.2	94.1	-18.2

measure were not in the hypothesized direction. The experimental group showed greater losses on the fluency measure than the control group did.

5. Grade Nine: The Mid- and Posttests. The fifth question asked in the study compared the writing changes of the grade-nine experimental and control students over the second three months of the study. The question asked was

> Do treatment classes of ninth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown also show significantly greater growth in the quality and complexity of compositions written three and six months following the administration of the treatment when the compositions of the two groups are judged by experienced raters and when they are examined for syntactic density, average T-unit length, and writing fluency?

For purposes of statistical analysis this question was stated in terms of null hypotheses examining the quality, syntactic density, average T-unit length, and fluency of the compositions.

Null Hypothesis 5a: Compositions written by treatment classes of ninth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be judged significantly different in quality from those written by non-treatment classes when the compositions written three months following the administration of the treatment are compared with those written six months following the administration of the treatment.

TABLE 85

Mid-post comparison on the quality measure for grade-nine subjects.

Source	df	F-ratio	probability	decision
Treatment	1	0.117	.74	N.S.
Treatment - Teacher	1	0.320	.58	N.S.
Treatment - Sex	1	1.199	.28	N.S.

As Table 85 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the quality of the compositions written by the ninth-grade experimental and control students was not significantly different when the compositions written three months following the administration of the treatment were compared with those written six months following the administration of the treatment and when the initial differences between the two groups were adjusted by the procedure of analysis of covariance. The very high probability of .74 indicates that measurable differences between the two groups were very slight.

As Table 86 shows, the changes in writing quality

TABLE 86

Mid-post comparison: average raw scores on the quality measure for grade-nine subjects.

	n	Midtest Raw Scores $\bar{x}$	σ	Posttest Raw Scores $\bar{x}$	σ	Posttest Minus Midtest
Experimental						
Boys	8	5.95	0.90	5.35	1.70	-.60
Girls	10	5.74	1.23	5.48	0.80	-.26
Combined	18	5.83	1.10	5.42	1.29	-.41
Control						
Boys	6	5.77	1.04	6.79	1.65	+1.02
Girls	12	6.04	1.25	5.10	1.60	-0.94
Combined	18	5.95	1.19	5.66	1.80	-0.29

TABLE 87

Mid-post comparison: average raw scores on the reading measure for the grade-nine sample used in the quality analysis.

	n	Midtest Raw Scores $\bar{x}$	σ	Posttest Raw Scores $\bar{x}$	σ	Posttest Minus Midtest
Experimental						
Boys	8	35.25	11.26	29.13	13.29	-6.12
Girls	10	38.50	13.43	42.10	12.81	+3.60
Combined	18	37.06	12.62	36.33	14.53	-0.73
Control						
Boys	6	44.33	12.71	46.50	11.80	+2.17
Girls	12	37.67	15.53	37.42	16.57	-0.25
Combined	18	39.89	14.98	40.44	15.74	+0.55

were not in the hypothesized direction. The minor differences in writing quality over the second three months favored the control groups. Table 87 presents the mid- and posttest reading scores of the grade-nine sample used

in the quality analysis.

> Null Hypothesis 5b: Compositions written by treatment classes of ninth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in syntactic density from those written by non-treatment classes when compositions written three months following the administration of the treatment are compared with those written six months following the administration of the treatment.

TABLE 88

Mid-post comparison on the syntactic density measure for grade-nine subjects.

Source	df	F-ratio	probability	decision
Treatment	1	.003	.96	N.S.
Treatment - Teacher	1	.049	.83	N.S.
Treatment - Sex	1	.586	.45	N.S.

As Table 88 shows, at the .05 level of confidence, the null hypothesis was not rejected and it is therefore concluded that the syntactic density of the compositions written by the ninth-grade experimental and control students was not significantly different when the compositions written three months following the administration of the treatment were compared with those written six months following the administration of the treatment and when initial differences between the two groups were adjusted by the procedure of analysis of covariance. The very high probability of .96 indicates that measurable differences between the two groups were indeed very slight.

TABLE 89

Mid-post comparison: average raw scores on the syntactic density measure for grade-nine subjects.

	n	Midtest Raw Scores $\bar{x}$	σ	Posttest Raw Scores $\bar{x}$	σ	Posttest Minus Midtest
Experimental						
Boys	11	3.73	0.67	3.85	0.69	+.12
Girls	13	3.71	0.83	3.27	0.68	-.44
Combined	24	3.72	0.76	3.54	0.74	-.18
Control						
Boys	9	3.69	1.17	3.34	1.07	-.35
Girls	15	3.65	0.86	3.60	0.86	-.05
Combined	24	3.67	0.99	3.50	0.96	-.17

TABLE 90

Mid-post comparison: average raw scores on the reading measure for the grade-nine sample used in the syntactic density analysis.

	n	Midtest Raw Scores $\bar{x}$	σ	Posttest Raw Scores $\bar{x}$	σ	Posttest Minus Midtest
Experimental						
Boys	11	30.73	13.44	25.55	13.18	-5.18
Girls	13	39.31	15.34	43.54	14.39	+4.23
Combined	24	35.38	15.12	35.29	16.50	-0.09
Control						
Boys	9	43.44	13.53	44.00	10.59	+0.56
Girls	15	40.27	16.39	39.93	18.22	-0.34
Combined	24	41.46	15.45	41.46	15.92	0

Table 89 shows that the changes in syntactic density were not in the hypothesized direction. Over the second three months of the study the experimental and control groups showed almost equal losses on the syntactic density measure.

Table 90 presents the mid- and posttest reading scores of the grade-nine sample used in the syntactic density analysis.

> Null Hypothesis 5c: Compositions written by treatment classes of ninth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in T-unit length from those written by non-treatment classes when the compositions written three months following the administration of the treatment are compared with those written six months following the administration of the treatment.

TABLE 91

Mid-post comparison on the T-unit measure for grade-nine subjects.

Source	df	F-ratio	probability	decision
Treatment	1	0.350	.56	N.S.
Treatment - Teacher	1	1.518	.22	N.S.
Treatment - Sex	1	0.232	.63	N.S.

As Table 91 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the average T-unit length of the compositions written by the ninth-grade experimental and control students was not significantly different when the compositions written three months following the administration of the treatment were compared with those written six months following the administration of the treatment and when initial differences between the two groups were adjusted by the procedure of analysis of covariance. The high probability of .56 suggests

that measurable differences between the two groups were small.

TABLE 92

Mid-post comparison: average raw scores on the T-unit measure for grade-nine subjects.

	n	Midtest Raw Scores $\bar{x}$	σ	Posttest Raw Scores $\bar{x}$	σ	Posttest Minus Midtest
Experimental						
Boys	28	14.38	2.77	14.68	2.32	+0.30
Girls	22	14.09	3.47	14.16	3.26	+0.07
Combined	50	14.25	3.10	14.45	2.78	+0.20
Control						
Boys	24	14.21	2.78	13.90	2.61	-0.31
Girls	29	14.16	2.58	15.59	3.24	+1.43
Combined	53	14.18	2.67	14.83	3.09	+0.65

TABLE 93

Mid-post comparison: average raw scores on the reading measure for the grade-nine sample used in the T-unit and fluency analyses.

	n	Midtest Raw Scores $\bar{x}$	σ	Posttest Raw Scores $\bar{x}$	σ	Posttest Minus Midtest
Experimental						
Boys	28	32.19	15.60	32.61	16.21	+0.42
Girls	22	37.82	14.70	43.86	13.84	+6.04
Combined	50	34.66	15.47	37.56	16.21	+2.90
Control						
Boys	24	42.75	14.37	44.38	13.29	+1.63
Girls	29	38.79	14.42	40.41	15.91	+1.62
Combined	53	40.59	14.53	42.21	14.91	+1.62

As Table 92 shows, the changes in average T-unit length over the second three months of the experiment were not in the hypothesized direction. Although both the

experimental and control students showed gains on this measure, the gains shown by the control students exceeded those shown by the experimental students. Table 93 presents the mid- and posttest reading scores of the grade-nine sample used in the T-unit analysis and the fluency analysis (below).

> Null Hypothesis 5d: Compositions written by treatment classes of ninth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in fluency from those written by the non-treatment classes when the compositions written three months following the administration of the treatment are compared with those written six months following the administration of the treatment.

TABLE 94

Mid-post comparison on the fluency measure for grade-nine subjects.

Source	df	F-ratio	probability	decision
Treatment	1	.724	.40	N.S.
Treatment - Teacher	1	.744	.39	N.S.
Treatment - Sex	1	.004	.95	N.S.

As Table 94 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the fluency of the compositions written by the ninth-grade experimental and control students was not significantly different when compositions written three months following the administration of the treatment were compared with those written six months following the administration of the treatment and when initial differences between the two

groups were adjusted by the procedure of analysis of covariance. The probability of .40 suggests that measurable differences between the two groups were small.

TABLE 95

Mid-post comparison: average raw scores on the fluency measure for grade-nine subjects.

	n	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Midtest
Experimental						
Boys	28	258.4	80.9	198.2	82.6	-60.2
Girls	22	265.3	106.5	232.2	96.2	-33.1
Combined	50	261.4	93.1	213.1	90.4	-48.3
Control						
Boys	24	250.0	110.2	208.9	65.6	-41.1
Girls	29	262.2	77.7	249.3	65.5	-12.9
Combined	53	256.6	94.0	231.0	68.5	-25.6

As Table 95 shows, the changes on the fluency measure over the second three months of the study were not in the hypothesized direction. Both the experimental and control groups showed losses over this period but the losses were greater for the experimental group than they were for the control group.

6. Grade Ten: The Mid- and Posttests. The sixth question asked in the study compared the writing changes of the grade-ten experimental and control classes over the second three-months of the study. The question asked was

> Do treatment classes of tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown also show

> significantly greater growth in the quality and complexity of compositions written three and six months following the administration of the treatment when the compositions of the two groups are judged by experienced raters and when they are examined for syntactic density, average T-unit length, and writing fluency?

For purposes of statistical analysis this question was stated in terms of null hypotheses examining the quality, syntactic density, average T-unit length, and writing fluency of the compositions.

> Null Hypothesis 6a: Compositions written by treatment classes of tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be judged significantly different in quality from those written by non-treatment classes when the compositions written three months following the administration of the treatment are compared with those written six months following the administration of the treatment.

TABLE 96

Mid-post comparison on the quality measure for grade-ten subjects.

Source	df	F-ratio	probability	decision
Treatment	1	0.165	.69	N.S.
Treatment - Teacher	1	1.017	.32	N.S.
Treatment - Sex	1	0.566	.46	N.S.

As Table 96 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the quality of the compositions written by the tenth-grade experimental and control students was not significantly different when the compositions written three

months following the administration of the treatment were compared with those written six months following the administration of the treatment and when initial differences between the two groups were adjusted by the procedure of analysis of covariance. The very high probability of .69 indicates that measurable differences between the two groups were slight.

TABLE 97

Mid-post comparison: average raw scores on the quality measure for grade-ten subjects.

	n	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Midtest
Experimental						
Boys	13	5.32	1.13	4.96	1.57	-0.36
Girls	5	6.57	1.02	6.23	1.23	-0.34
Combined	18	5.66	1.24	5.31	1.60	-0.35
Control						
Boys	6	6.08	2.24	4.65	0.79	-1.43
Girls	12	6.33	2.44	6.37	1.66	+0.04
Combined	18	6.25	2.38	5.79	1.64	-0.46

As Table 97 shows, the changes in writing quality were in the hypothesized direction. Minor differences in writing quality over the second three months favored the experimental groups. Table 98 presents the mid- and posttest reading scores of the grade-ten samples used in the quality analysis.

> Null Hypothesis 6b: Compositions written by treatment classes of tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly

TABLE 98

Mid-post comparison: average raw scores on the reading measure for the grade-ten sample used in the quality analysis.

	n	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Midtest
Experimental						
Boys	13	53.46	15.18	49.46	14.39	-4.00
Girls	5	43.80	12.92	40.20	17.65	-3.60
Combined	18	50.78	15.22	46.89	15.91	-3.89
Control						
Boys	6	48.83	13.43	47.67	20.67	-1.16
Girls	12	53.42	18.91	51.33	17.49	-2.09
Combined	18	51.89	17.42	50.11	18.69	-1.78

different in syntactic density from those written by non-treatment classes when the compositions written three months following the administration of the treatment are compared with those written six months following the administration of the treatment.

TABLE 99

Mid-post comparison on the syntactic density measure for the grade-ten subjects.

Source	df	F-ratio	probability	decision
Treatment	1	2.860	.10	N.S.
Treatment - Teacher	1	0.282	.60	N.S.
Treatment - Sex	1	0.432	.52	N.S.

As Table 99 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the syntactic density of the compositions

written by the tenth-grade experimental and control students was not significantly different when the compositions written three months following the administration of the treatment were compared with those written six months following the administration of the treatment and when the initial differences between the two groups were adjusted by the procedure of analysis of covariance. The probability of .10 indicates measurable differences between the two groups were not substantial.

TABLE 100

Mid-post comparison: average raw scores on the syntactic density measure for the grade-ten subjects.

		Midtest Raw Scores		Posttest Raw Scores		Pretest Minus Midtest
	n	$\bar{x}$	σ	$\bar{x}$	σ	
Experimental						
Boys	16	3.90	0.87	3.90	.87	0
Girls	8	3.57	0.96	3.80	.60	+.23
Combined	24	3.79	0.92	3.87	.79	+.08
Control						
Boys	9	4.43	0.73	3.67	.72	-.76
Girls	15	4.05	1.04	4.00	.92	-.05
Combined	24	4.19	0.95	3.88	.87	-.31

Table 100 shows that the changes in syntactic density were in the hypothesized direction. Over the second three months of the experiment the experimental group showed slight gains while the control groups showed losses. Table 101 presents the mid- and posttest reading scores of the grade-ten sample used in the syntactic density analysis.

TABLE 101

Mid-post comparison: average raw scores on the reading measure for the grade-ten sample used in the syntactic density analysis.

	n	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Midtest
Experimental						
Boys	16	52.81	13.75	49.75	13.33	-3.06
Girls	8	49.38	12.66	46.50	16.39	-2.88
Combined	24	51.67	13.50	48.67	14.50	-3.00
Control						
Boys	9	51.11	12.97	49.56	18.99	-1.55
Girls	15	54.00	17.36	52.87	16.02	-1.13
Combined	24	52.92	15.92	51.63	17.27	-1.29

Null Hypothesis 6c: Compositions written by treatment classes of tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in T-unit length from those written by non-treatment classes when the compositions written three months following the administration of the treatment are compared with those written six months following the administration of the treatment.

TABLE 102

Mid-post comparison on the T-unit measure for the grade-ten subjects.

Source	df	F-ratio	probability	decision
Treatment	1	2.146	.15	N.S.
Treatment - Teacher	1	2.736	.10	N.S.
Treatment - Sex	1	0.0001	.99	N.S.

As Table 102 shows, at the .05 level of confidence

the null hypothesis was not rejected and it is therefore concluded that the average T-unit length of the compositions written by the tenth-grade experimental and control students was not significantly different when the compositions written three months following the administration of the treatment were compared with those written six months following the administration of the treatment and when initial differences between the two groups were adjusted by the procedure of analysis of covariance. The probability of .15 suggests that measurable differences between the two groups were not substantial.

TABLE 103

Mid-post comparison: average raw scores on the T-unit measure for grade-ten subjects.

	n	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Midtest
Experimental						
Boys	25	14.96	2.99	14.73	2.99	-0.23
Girls	19	15.47	3.24	15.33	3.65	-0.14
Combined	44	15.18	3.11	14.99	3.31	-0.19
Control						
Boys	23	15.97	3.02	15.10	3.29	-0.87
Girls	24	15.75	3.54	14.72	3.44	-1.03
Combined	47	15.86	3.30	14.91	3.37	-0.95

As Table 103 shows, the changes in average T-unit length over the second three months of the study were in the hypothesized direction. Although both the experimental and control groups showed losses during this period, the control

TABLE 104

Mid-post comparison: average raw scores on the reading measure for the grade-ten sample used in the T-unit and fluency analyses.

	n	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Midtest
Experimental						
Boys	25	54.48	12.85	51.96	11.93	-2.52
Girls	19	55.26	14.83	52.26	18.47	-3.00
Combined	44	54.82	13.74	52.09	15.10	-2.73
Control						
Boys	23	51.22	12.12	48.48	16.41	-2.74
Girls	24	55.21	17.84	55.08	16.51	-0.13
Combined	47	53.26	15.44	51.85	16.79	-1.41

group showed greater losses than the experimental group did. Table 104 presents the mid- and posttest reading scores of the grade-ten samples used in the T-unit analysis and the fluency analysis (below).

Null Hypothesis 6d: Compositions written by treatment classes of tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in fluency from those written by non-treatment classes when the compositions written three months following the administration of the treatment are compared with those written six months following the administration of the treatment.

As Table 105 shows, at the .05 level of confidence the null hypothesis was rejected and it is therefore concluded that the fluency of the compositions written by the tenth-grade experimental and control students was significantly different when the compositions written three months following the administration of the treatment were compared with those written six months following the administration

TABLE 105

Mid-post comparison on the fluency measure for the grade-ten subjects.

Source	df	F-ratio	probability	decision
Treatment	1	4.942	.03	Sig
Treatment - Teacher	1	0.432	.51	N.S.
Treatment - Sex	1	4.846	.03	Sig

of the treatment and when the initial differences between the two groups were adjusted by the procedure of analysis of covariance. The changes in fluency, however, favored the control groups.

Table 105 also shows that at the .05 level of confidence there was a significant interaction between the treatment and the sex of the subject.

TABLE 106

Mid-post comparison: average raw scores on the fluency measure for grade-ten subjects.

	n	Midtest Raw Scores $\bar{x}$	Midtest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Midtest
Experimental						
Boys	25	286.7	72.4	255.4	93.5	-31.3
Girls	19	318.4	79.3	257.9	72.8	-60.5
Combined	44	300.4	77.0	256.5	85.2	-43.9
Control						
Boys	23	290.4	58.4	274.4	107.3	-16.0
Girls	24	244.1	121.9	240.2	119.2	-3.9
Combined	47	266.8	98.9	256.9	114.8	-9.9

As Table 106 shows, the changes in fluency were not in the hypothesized direction. Although both the experimental and control groups showed losses in fluency scores over the second three months of the experiment, these losses were significantly greater for the experimental groups than for the control groups. As was noted in the discussion of Null Hypothesis 3d, the experimental students showed considerable gains in fluency scores over the first three months of the experiment while the control students showed losses. Over the second three months, however, the experimental students showed heavy losses while the control students again showed moderate losses. Over the six months of the experiment (cf. Null Hypothesis 9d) the experimental and control students showed equal losses, but since the scores of the experimental students fluctuated so widely, the differences on the mid- and posttest comparison were significant in favor of the control groups. These fluctuations were also responsible for the significant interaction between the treatment and the sex of the subject. The control girls had consistent fluency scores on all three tests. The experimental girls, on the other hand, showed gains on the midtest but losses on the posttest, resulting in considerable losses when the mid- and posttests were compared.

7. Grades Nine and Ten: Pre- and Posttests. The seventh question asked in the study compared the writing changes of the grade-nine and grade-ten experimental and control students over the full six months of the study. The question

asked was

> Do treatment classes of ninth- and tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown also show significantly greater growth in the quality and complexity of compositions written six months following the administration of the treatment when the compositions of the two groups are judged by experienced raters and when they are examined for syntactic density, T-unit length, and writing fluency?

For purposes of statistical analysis this question was stated in terms of null hypotheses examining the quality, syntactic density, average T-unit length, and writing fluency of the compositions.

> Null Hypothesis 7a: Compositions written by treatment classes of ninth- and tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be judged significantly different in quality from those written by non-treatment classes when the compositions are written six months following the administration of the treatment.

TABLE 107

Pre-post comparison on the quality measure for grades nine and ten combined.

Source	df	F-ratio	probability	decision
Treatment	1	.030	.86	N.S.
Treatment - Teacher	3	.856	.47	N.S.
Treatment - Sex	1	.006	.95	N.S.

As Table 107 shows, at the .05 level of confidence, the null hypothesis was not rejected and it is therefore concluded that the quality of the compositions written by the ninth- and tenth-grade experimental and control students

was not significantly different six months following the administration of the treatment when initial differences between the two groups were adjusted by the procedure of analysis of covariance. The very high probability of .86 indicates that measurable differences between the two groups were indeed slight.

TABLE 108

Pre-post comparison: average raw scores on the quality measure for grades nine and ten combined.

	n	Pretest Raw Scores $\bar{x}$	σ	Posttest Raw Scores $\bar{x}$	σ	Posttest Minus Pretest
Experimental						
Boys	21	5.81	1.29	5.11	1.63	-.70
Girls	15	5.71	1.17	5.73	1.03	+.02
Combined	36	5.77	1.24	5.37	1.45	-.40
Control						
Boys	12	6.10	1.68	5.72	1.68	-.38
Girls	24	5.63	1.63	5.74	1.75	+.11
Combined	36	5.79	1.66	5.73	1.72	-.06

As Table 108 shows, the changes in writing quality over the six months of the experiment were not in the hypothesized direction. Slight differences in quality favored the control group. Table 109 presents the pre- and posttest reading scores of the ninth- and tenth-grade samples used in the quality analysis.

> Null Hypothesis 7b: Compositions written by treatment classes of ninth- and tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in syntactic density from those written by non-treatment classes when the compositions are written six months following the administration of the treatment.

TABLE 109

Pre-post comparison: average raw scores on the reading measure for the grade-nine and -ten samples used in the quality analysis.

	n	Pretest Raw Scores $\bar{x}$	σ	Posttest Raw Scores $\bar{x}$	σ	Posttest Minus Pretest
Experimental						
Boys	21	41.52	15.21	41.71	17.12	+0.19
Girls	15	41.00	13.47	41.47	14.63	+0.47
Combined	36	41.31	14.51	41.61	16.13	+0.30
Control						
Boys	12	47.25	13.14	47.08	16.84	-0.17
Girls	24	47.67	15.82	44.38	18.40	-3.29
Combined	36	47.53	14.98	45.28	17.94	-2.25

TABLE 110

Pre-post comparison on the syntactic density measure for grades nine and ten combined.

Source	df	F-ratio	probability	decision
Treatment	1	1.131	.26	N.S.
Treatment - Teacher	3	0.680	.57	N.S.
Treatment - Sex	1	1.139	.24	N.S.

As Table 110 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the syntactic density of the compositions written by the ninth- and tenth-grade experimental and control students was not significantly different six months following the administration of the treatment when initial differences

between the two groups were adjusted by the procedure of analysis of covariance. The probability of .26 indicates that measurable differences between the two groups were small.

TABLE 111

Pre-post comparison: average raw scores on the syntactic density measure for grades nine and ten combined.

	n	Pretest Raw Scores $\bar{x}$	σ	Posttest Raw Scores $\bar{x}$	σ	Posttest Minus Pretest
Experimental						
Boys	27	3.62	0.83	3.88	.80	+.26
Girls	21	3.51	0.75	3.48	.70	-.03
Combined	48	3.57	0.80	3.70	.79	+.13
Control						
Boys	18	3.94	1.18	3.50	.93	-.44
Girls	30	3.62	0.90	3.80	.91	+.18
Combined	48	3.74	1.03	3.69	.93	-.05

TABLE 112

Pre-post comparison: average raw scores on the reading measure for the grade-nine and -ten samples used in the syntactic density analysis.

	n	Pretest Raw Scores $\bar{x}$	σ	Posttest Raw Scores $\bar{x}$	σ	Posttest Minus Pretest
Experimental						
Boys	27	40.11	16.34	39.89	17.82	-0.22
Girls	21	44.48	14.00	44.68	15.25	+0.20
Combined	48	42.02	15.51	41.98	16.91	-0.04
Control						
Boys	18	47.39	13.65	46.78	15.62	-0.61
Girls	30	49.70	15.57	46.40	18.33	-3.30
Combined	48	48.83	14.92	46.54	17.37	-2.29

As Table 111 shows, the changes on the syntactic density measure were in the hypothesized direction. Over the six months of the experiment, the experimental group showed small gains on the syntactic density measure while the control group showed slight losses. Table 112 presents the pre- and posttest reading scores of the ninth- and tenth-grade samples used in the syntactic density analysis.

> Null Hypothesis 7c: Compositions written by treatment classes of ninth- and tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in average T-unit length from those written by non-treatment classes when the compositions are written six months following the administration of the treatment.

TABLE 113

Pre-post comparison on the T-unit measure for grades nine and ten combined.

Source	df	F-ratio	probability	decision
Treatment	1	0.478	.49	N.S.
Treatment - Teacher	3	1.198	.31	N.S.
Treatment - Sex	1	4.384	.04	Sig

As Table 113 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the average T-unit length of the compositions written by the ninth- and tenth-grade experimental and control students was not significantly different six months following the administration of the treatment when initial

differences between the two groups were adjusted by the procedure of analysis of covariance. The high probability of .49 indicates that measurable differences between the two groups were small.

Table 113 also shows that there was a significant interaction at the .04 level of confidence between the treatment and the sex of the subject.

TABLE 114

Pre-post comparison: average raw scores on the T-unit measure for grades nine and ten combined.

	n	Pretest Raw Scores $\bar{x}$	σ	Posttest Raw Scores $\bar{x}$	σ	Posttest Minus Pretest
Experimental						
Boys	53	14.44	3.06	14.70	2.66	+.26
Girls	41	15.13	3.06	14.71	3.49	-.42
Combined	94	14.74	3.08	14.70	3.05	-.04
Control						
Boys	47	15.00	3.12	14.49	3.02	-.51
Girls	53	14.39	2.80	15.20	3.36	+.81
Combined	100	14.68	2.97	14.86	3.22	+.18

As Table 114 shows, the changes in average T-unit length were not in the hypothesized direction. Over the six months of the experiment the control students showed slight gains on this measure while the experimental students showed losses. During this time the experimental boys and the control girls showed gains while the experimental girls and the control boys showed losses. This combination of gains and losses by the boys and girls accounts for the interaction between the treatment and the sex of the subject, an

TABLE 115

Pre-post comparison: average raw scores on the reading measure for the grade-nine and -ten samples used in the T-unit and fluency analyses.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Pretest
Experimental						
Boys	53	39.60	16.36	41.74	17.30	+2.14
Girls	41	45.44	16.36	47.76	16.69	+2.32
Combined	94	42.15	16.62	44.36	17.29	+2.21
Control						
Boys	47	47.11	14.05	46.38	15.04	-0.73
Girls	53	47.74	16.46	47.06	17.75	-0.68
Combined	100	47.44	15.37	46.74	16.54	-0.70

interaction which must be regarded as spurious in light of the large number of interactions tested and the few which proved to be statistically significant. Table 115 presents the pre- and posttest reading scores of the ninth- and tenth-grade samples used in the T-unit analysis and the fluency analysis (below).

> Null Hypothesis 7d: Compositions written by treatment classes of ninth- and tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in fluency from those written by non-treatment classes when the compositions are written six months following the administration of the treatment.

As Table 116 indicates, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the fluency of the compositions written by the ninth- and tenth-grade experimental and control students was not significantly different six months following the administration of the treatment when initial

TABLE 116

Pre-post comparison on the fluency measure for grades nine and ten combined.

Source	df	F-ratio	probability	decision
Treatment	1	0.0003	.99	N.S.
Treatment - Teacher	3	0.493	.69	N.S.
Treatment - Sex	1	2.255	.14	N.S.

differences between the two groups were adjusted by the procedure of analysis of covariance. The very high probability of .99 indicates that measurable differences between the two groups were indeed very slight.

TABLE 117

Pre-post comparison: average raw scores on the fluency measure for grades nine and ten combined.

	n	Pretest Raw Scores $\bar{x}$	σ	Posttest Raw Scores $\bar{x}$	σ	Posttest Minus Pretest
Experimental						
Boys	53	252.3	100.5	225.2	92.4	-27.1
Girls	41	273.8	100.1	244.1	87.1	-29.7
Combined	94	261.7	100.9	233.4	90.6	-28.3
Control						
Boys	47	280.8	95.2	241.0	94.4	-39.8
Girls	53	261.4	98.2	245.2	93.8	-16.2
Combined	100	270.5	97.3	243.2	94.1	-27.3

Table 117 shows that changes on the fluency measure were not in the hypothesized direction. Over the six months

of the experiment the experimental and control groups showed very similar losses in fluency with slight differences favoring the control group.

8. Grade Nine: The Pre- and Posttests. The eighth question asked in the study compared the writing changes of the grade-nine experimental and control students over the six months of the study. The question asked was

> Do treatment classes of ninth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown also show significantly greater growth in the quality and complexity of compositions written six months following the administration of the treatment when the compositions of the two groups are judged by experienced raters and when they are examined for syntactic density, T-unit length, and writing fluency?

For purposes of statistical analysis this question was stated in terms of null hypotheses examining the quality, syntactic density, average T-unit length, and fluency of the compositions.

> Null Hypothesis 8a: Compositions written by treatment classes of ninth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be judged significantly different in quality from those written by non-treatment classes when the compositions are written six months following the administration of the treatment.

As Table 118 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the quality of the compositions written by the ninth-grade experimental and control students was not significantly different six months following the administration of the treatment when initial differences between the two groups were adjusted by the procedure of analysis of covariance. The

TABLE 118

Pre-post comparison on the quality measure for grade-nine subjects.

Source	df	F-ratio	probability	decision
Treatment	1	0.593	.45	N.S.
Treatment - Teacher	1	0.404	.53	N.S.
Treatment - Sex	1	1.000	.33	N.S.

high probability of .45 suggests that measurable differences between the two groups were small.

TABLE 119

Pre-post comparison: average raw scores on the quality measure for grade-nine subjects.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Pretest
Experimental						
Boys	8	6.16	1.32	5.35	1.70	-0.81
Girls	10	5.22	1.72	5.48	0.80	+0.26
Combined	18	5.64	1.13	5.42	1.29	-0.22
Control						
Boys	6	5.64	2.15	6.79	1.65	+1.15
Girls	12	5.30	1.72	5.10	1.60	-0.20
Combined	18	5.41	1.88	5.66	1.80	+0.25

Table 119 shows that differences on the quality measure were not in the hypothesized direction. Over the six months of the experiment the control group showed small gains in writing quality while the experimental group showed small

TABLE 120

Pre-post comparison: average raw scores on the reading measure for the grade-nine sample used in the quality analysis.

	n	Pretest Raw Scores $\bar{x}$	σ	Posttest Raw Scores $\bar{x}$	σ	Posttest Minus Pretest
Experimental						
Boys	8	33.38	11.77	29.13	13.29	-4.25
Girls	10	39.90	12.72	42.10	12.81	+2.20
Combined	18	37.00	12.73	36.33	14.53	-0.67
Control						
Boys	6	46.67	12.26	46.50	11.80	-0.17
Girls	12	41.42	13.43	37.42	16.57	-4.00
Combined	18	43.17	13.28	40.44	15.74	-2.73

losses. Table 120 presents the pre- and posttest reading scores of the grade-nine sample used in the quality analysis.

> Null Hypothesis 8b: Compositions written by treatment classes of ninth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in syntactic density from those written by the non-treatment classes when the compositions are written three months following the administration of the treatment.

TABLE 121

Pre-post comparison on the syntactic density measure for grade-nine subjects.

Source	df	F-ratio	probability	decision
Treatment	1	0.704	.41	N.S.
Treatment - Teacher	1	1.162	.21	N.S.
Treatment - Sex	1	0.003	.96	N.S.

As Table 121 indicates, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the syntactic density of the compositions written by the ninth-grade experimental and control students was not significantly different six months following the administration of the treatment when initial differences between the two groups were adjusted by the procedure of analysis of covariance. The high probability of .41 indicates that measurable differences between the two groups were small.

TABLE 122

Pre-post comparison: average raw scores on the syntactic density measure for grade-nine subjects.

	n	Pretest Raw Scores $\bar{x}$	σ	Posttest Raw Scores $\bar{x}$	σ	Posttest Minus Pretest
Experimental						
Boys	11	3.34	.61	3.85	0.69	+.51
Girls	13	3.64	.69	3.27	0.68	-.37
Combined	24	3.51	.67	3.54	0.74	+.03
Control						
Boys	9	3.45	.45	3.34	1.07	-.11
Girls	15	3.86	.83	3.60	0.86	-.26
Combined	24	3.70	.74	3.50	0.96	-.20

As Table 122 shows, the changes in syntactic density were in the hypothesized direction. Over the six months of the study very small differences in syntactic density favored the experimental group. Table 123 presents the pre- and posttest reading scores of the ninth-grade sample used in the syntactic density analysis.

TABLE 123

Pre-post comparison: average raw scores on the reading measure for the grade-nine sample used in the syntactic density analysis.

	n	Pretest Raw Scores $\bar{x}$	σ	Posttest Raw Scores $\bar{x}$	σ	Posttest Minus Pretest
Experimental						
Boys	11	28.82	12.90	25.55	13.18	-3.27
Girls	13	42.69	14.37	43.54	14.39	+0.85
Combined	24	36.33	15.36	35.29	16.50	-1.04
Control						
Boys	9	44.56	12.93	44.00	10.59	-0.56
Girls	15	44.67	14.81	39.93	18.22	-4.74
Combined	24	44.63	14.13	41.46	15.92	-3.17

Null Hypothesis 8c: Compositions written by treatment classes of ninth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in average T-unit length from those written by non-treatment classes when the compositions are written six months following the administration of the treatment.

TABLE 124

Pre-post comparison on the T-unit measure for grade-nine subjects.

Source	df	F-ratio	probability	decision
Treatment	1	.286	.59	N.S.
Treatment - Teacher	1	.212	.65	N.S.
Treatment - Sex	1	.001	.97	N.S.

As Table 124 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore

concluded that the average T-unit length of the compositions written by the ninth-grade experimental and control students was not significantly different three months following the administration of the treatment when initial differences between the two groups were adjusted by the procedure of analysis of covariance. The high probability of .59 suggests that measurable differences between the two groups were small.

TABLE 125

Pre-post comparison: average raw scores on the T-unit measure for grade-nine subjects.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Pretest
Experimental						
Boys	28	13.73	2.92	14.68	2.32	+.95
Girls	22	14.29	2.70	14.16	2.78	-.13
Combined	50	13.97	2.84	14.45	2.78	+.48
Control						
Boys	24	13.93	2.23	13.90	2.61	-.03
Girls	29	15.00	2.95	15.59	3.24	+.59
Combined	53	14.51	2.70	14.83	3.09	+.32

Table 125 shows that the differences in average T-unit length were in the hypothesized direction. Over the six months of the study slight differences favored the experimental group. Table 126 presents the pre- and posttest reading scores of the ninth-grade sample used in the T-unit analysis and the fluency analysis (below).

> Null Hypothesis 8d: Compositions written by treatment classes of ninth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in fluency from those written by non-treatment classes when the compositions are written six months

TABLE 126

Pre-post comparison: average raw scores on the reading measure for the grade-nine sample used in the T-unit and fluency analysis.

	n	Pretest Raw Scores $\bar{x}$	σ	Posttest Raw Scores $\bar{x}$	σ	Posttest Minus Pretest
Experimental						
Boys	28	31.79	15.47	32.61	16.21	+0.82
Girls	22	40.41	13.28	43.86	13.84	+3.45
Combined	50	35.58	15.16	37.56	16.21	+1.98
Control						
Boys	24	44.25	14.47	44.38	13.29	+0.13
Girls	29	41.45	14.27	40.41	15.91	-1.04
Combined	53	42.72	14.43	42.21	14.91	-0.51

following the administration of the treatment.

TABLE 127

Pre-post comparison on the fluency measure for grade-nine subjects.

Source	df	F-ratio	probability	decision
Treatment	1	.668	.42	N.S.
Treatment - Teacher	1	.003	.92	N.S.
Treatment - Sex	1	.003	.96	N.S.

As Table 127 indicates, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the fluency of the compositions written by the ninth-grade experimental and control students was not significantly different three months following the

administration of the treatment when initial differences between the two groups were adjusted by the procedure of analysis of covariance. The high probability of .42 suggests that measurable differences between the two groups were small.

TABLE 128

Pre-post comparison: average raw scores on the fluency measure for grade-nine subjects.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Pretest
Experimental						
Boys	28	241.3	91.4	198.2	82.6	-43.1
Girls	22	250.7	102.3	232.2	96.2	-18.5
Combined	50	245.4	96.5	213.1	90.4	-32.3
Control						
Boys	24	256.7	71.6	208.9	65.6	-47.8
Girls	29	275.4	86.0	249.3	65.5	-26.1
Combined	53	266.9	80.3	231.0	68.5	-35.9

As Table 128 shows, the changes on the fluency measure were in the hypothesized direction. Over the six months of the study very small differences on the fluency measure favored the experimental group.

9. Grade Ten: Pre- and Posttests. The ninth question asked in the study compared the writing changes of the grade-ten experimental and control students over the six months of the study. The question asked was

> Do treatment classes of tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown also show significantly greater growth in the quality and complexity of compositions written six months following

> the administration of the treatment when the compositions of the two groups are judged by experienced raters and when they are examined for syntactic density, average T-unit length, and writing fluency?

For purposes of statistical analysis this question was stated in terms of null hypotheses examining the quality, syntactic density, average T-unit length, and fluency of the compositions.

> Null Hypothesis 9a: Compositions written by treatment classes of tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be judged significantly different in quality from those written by non-treatment classes when the compositions are written three months following the administration of the treatment.

TABLE 129

Pre-post comparison on the quality measure for grade-ten subjects.

Source	df	F-ratio	probability	decision
Treatment	1	1.658	.21	N.S.
Treatment - Teacher	1	0.000	.99	N.S.
Treatment - Sex	1	5.733	.02	Sig

As Table 129 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the quality of the compositions written by the tenth-grade experimental and control students was not significantly different six months following the administration of the treatment when initial differences

between the two groups were adjusted by the procedure of analysis of covariance. The probability of .21 indicates that measurable differences between the two groups were small.

Table 129 also shows that at the .05 level of confidence there was a significant interaction between the treatment and the sex of the subject.

TABLE 130

Pre-post comparison: average raw scores on the quality measure for grade-ten subjects.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Pretest
Experimental						
Boys	13	5.60	1.23	4.96	1.57	-0.64
Girls	5	6.68	1.30	6.23	1.23	-0.45
Combined	18	5.90	1.34	5.31	1.60	-0.59
Control						
Boys	6	6.57	0.76	4.65	0.79	-1.92
Girls	12	5.96	1.47	6.37	1.66	+0.41
Combined	18	6.16	1.31	5.79	1.64	-0.37

Table 130 shows that differences on the quality measure were not in the hypothesized direction. Although both the experimental and control groups showed losses on the quality measure over the six months of the experiment, the control group showed fewer losses than the experimental group did. Table 130 also shows the reason for the significant interaction between the treatment and the sex of the subject noted above. The control girls showed gains on the quality measure while the control boys showed substantial losses. At the same time, the experimental boys

TABLE 131

Pre-post comparison: average raw scores on the reading measure for the grade-ten sample used in the quality analysis.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Pretest
Experimental						
Boys	13	46.54	14.91	49.46	14.39	+2.92
Girls	5	43.20	14.59	40.20	17.65	-3.00
Combined	18	45.61	13.90	46.89	15.91	+1.28
Control						
Boys	6	47.83	13.95	47.67	20.67	-0.16
Girls	12	53.92	15.56	51.33	17.49	-2.59
Combined	18	51.89	15.32	50.11	18.69	-1.78

and girls showed losses similar to each other. This pattern will be observed in both the T-unit length (Null Hypothesis 9c) and the fluency (Null Hypothesis 9d) which follow. Table 91 presents the pre- and posttest reading scores of the grade-ten sample used in the quality analysis.

> Null Hypothesis 9b: Compositions written by treatment classes of tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in syntactic density from those written by non-treatment classes when the compositions are written six months following the administration of the treatment.

As Table 132 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the syntactic density of the compositions written by the tenth-grade experimental and control students was not significantly different six months following the administration of the treatment when initial differences between the two groups were adjusted by the procedure of analysis of

TABLE 132

Pre-post comparison on the syntactic density measure for grade-ten subjects.

Source	df	F-ratio	probability	decision
Treatment	1	0.672	.42	N.S.
Treatment - Teacher	1	0.345	.56	N.S.
Treatment - Sex	1	1.632	.21	N.S.

covariance. The probability of .42 suggests that measurable differences between the two groups were small.

TABLE 133

Pre-post comparison: average raw scores on the syntactic density measure for the grade-ten subjects.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Pretest
Experimental						
Boys	16	3.81	0.90	3.90	.87	+.09
Girls	8	3.29	0.79	3.80	.60	+.51
Combined	24	3.64	0.90	3.87	.79	+.23
Control						
Boys	9	4.44	1.45	3.67	.72	-.77
Girls	15	3.38	0.91	4.00	.92	+.62
Combined	24	3.78	1.25	3.88	.87	+.10

Table 133 shows that differences on the syntactic density measure, although small and statistically non-significant, were in the hypothesized direction. Over the six months of the experiment the experimental group showed

TABLE 134

Pre-post comparison: average raw scores on the reading measure for the grade-ten sample used in the syntactic density analysis.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Pretest
Experimental						
Boys	16	47.88	13.72	49.75	13.33	+1.87
Girls	8	47.38	12.85	46.50	16.39	-0.88
Combined	24	47.71	13.44	48.67	14.50	+0.96
Control						
Boys	9	50.22	13.76	49.56	18.99	-0.66
Girls	15	54.73	14.66	52.87	16.02	-1.86
Combined	24	53.04	14.50	51.63	17.27	-1.41

slightly larger gains than the control group did. Table 134 presents the pre- and posttest reading scores of the grade-ten sample used in the syntactic density analysis.

> Null Hypothesis 9c: Compositions written by treatment classes of tenth-grade students who have shown significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in average T-unit length from those written by non-treatment classes when the compositions are written six months following the administration of the treatment.

As Table 135 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the average T-unit length of the compositions written by the tenth-grade experimental and control students was not significantly different six months following the administration of the treatment when initial differences between the two groups were adjusted by the procedure of analysis of covariance. The probability of .14 indicates

TABLE 135

Pre-post comparison on the T-unit measure for the grade-ten subjects.

Source	df	F-ratio	probability	decision
Treatment	1	2.202	.14	N.S.
Treatment - Teacher	1	1.078	.30	N.S.
Treatment - Sex	1	6.132	.02	Sig

that measurable differences between the two groups were not substantial.

Table 135 also shows that at the .05 level of confidence there was a significant interaction between the treatment and the sex of the subject.

TABLE 136

Pre-post comparison: average raw scores on the T-unit measure for grade-ten subjects.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Pretest
Experimental						
Boys	25	15.24	3.02	14.73	2.99	-0.51
Girls	19	16.11	3.16	15.33	3.65	-0.78
Combined	44	15.62	3.11	14.99	3.31	-0.63
Control						
Boys	23	16.13	3.51	15.10	3.29	-1.03
Girls	24	13.66	2.42	14.72	3.44	+1.06
Combined	47	14.87	3.24	14.91	3.37	+0.04

Table 136 shows that differences on the syntactic

TABLE 137

Pre-post comparison: average raw scores on the reading measure for the grade-ten sample used in the T-unit and fluency analyses.

	n	Pretest Raw Scores $\bar{x}$	σ	Posttest Raw Scores $\bar{x}$	σ	Posttest Minus Pretest
Experimental						
Boys	25	48.36	12.44	51.96	11.93	+3.60
Girls	19	51.26	17.61	52.26	18.47	+1.00
Combined	44	49.61	14.96	52.09	15.10	+2.48
Control						
Boys	23	50.09	12.93	48.48	16.41	-1.61
Girls	24	55.33	15.70	55.08	16.51	-0.25
Combined	47	52.77	14.65	51.85	16.79	-0.92

density measure were not in the hypothesized direction. Over the six months of the experiment the control group showed slight gains while the experimental group showed losses on the syntactic density measure. Table 136 also shows the reason for the significant interaction between the treatment and the sex of the subject. The control girls, who scored well below any other subgroup on the pretest, were the only subgroup to show gains in average T-unit length on the posttest. Despite these gains, the control girls still scored the lowest of any subgroup on the posttest. Nonetheless, these gains were responsible for significant interaction between the treatment and the sex of the subject. Table 137 presents the pre- and posttest reading scores of the grade-ten sample used in the T-unit and fluency analysis (below).

<u>Null Hypothesis 9d</u>: Compositions written by treatment classes of tenth-grade students who have shown

significantly greater gains on reading measures than non-treatment classes have shown will not be significantly different in fluency from those written by non-treatment classes when the compositions are written six months following the administration of the treatment.

TABLE 138

Pre-post comparison on the fluency measure for the grade-ten subjects.

Source	df	F-ratio	probability	decision
Treatment	1	0.126	.72	N.S.
Treatment - Teacher	1	0.307	.58	N.S.
Treatment - Sex	1	4.555	.04	Sig

As Table 138 shows, at the .05 level of confidence the null hypothesis was not rejected and it is therefore concluded that the fluency of the compositions written by the tenth-grade experimental and control students was not significantly different six months following the administration of the treatment when initial differences between the two groups were adjusted by the procedure of analysis of covariance. The high probability of .72 suggests that measurable differences between the experimental and control groups were slight.

Table 138 also shows that at the .05 level of confidence there was a significant interaction between the treatment and the sex of the subject.

Table 139 shows that differences on the fluency measure were not in the hypothesized direction. Over the six

TABLE 139

Pre-post comparison: average raw scores on the fluency measure for grade-ten subjects.

	n	Pretest Raw Scores $\bar{x}$	Pretest Raw Scores σ	Posttest Raw Scores $\bar{x}$	Posttest Raw Scores σ	Posttest Minus Pretest
Experimental						
Boys	25	264.7	108.4	255.4	93.5	-9.3
Girls	19	300.4	90.5	257.9	72.8	-42.5
Combined	44	280.1	102.6	256.5	85.2	-23.6
Control						
Boys	23	305.9	109.2	274.4	107.3	-31.5
Girls	24	244.5	108.8	240.2	119.2	-4.3
Combined	47	274.6	113.2	256.9	114.8	-17.7

months of the experiment small differences on the fluency measure favored the control group. Table 139 also shows the reason for the significant interaction between the treatment and the sex of the subject. The two subgroups who scored lowest on the pretest, the experimental boys and the control girls, showed the smallest losses over the six months of the experiment. These small losses combined with large losses by the experimental girls and control boys to produce a spurious interaction between the treatment and the sex of the subject.

B. Summary

Table 140 offers a summary of the statistical analyses of the 36 null hypotheses reported above. As Table 140 shows, only 3 of the 36 null hypotheses tested in the study were significant at the .05 level of confidence. All 3 of these tests favored the control groups. Of the 33 tests that

TABLE 140

Summary of Null Hypotheses

Questions	Null Hypotheses	Prob	Treatment Sig. at .05	Group Showing Greatest Gains	Teacher Interaction Significant at .05	Sex Interaction Significant at .05	Group Showing Greatest Gains
1	Gr 9 & 10 Pre-Mid						
	A Quality	0.97	N.S.	Control	N.S.	N.S.	
	B Syn Dens	0.49	N.S.	Control	N.S.	N.S.	
	C T-unit	0.32	N.S.	Control	N.S.	Sig.	Con Girls
	D Fluency	0.19	N.S.	Exper.	N.S.	N.S.	
2	Gr 9 Pre-Mid						
	A Quality	0.57	N.S.	Control	N.S.	N.S.	
	B Syn Dens	0.45	N.S.	Exper.	N.S.	N.S.	
	C T-unit	0.43	N.S.	Exper.	N.S.	N.S.	
	D Fluency	0.12	N.S.	Exper.	N.S.	N.S.	
3	Gr 10 Pre-Mid						
	A Quality	0.60	N.S.	Control	N.S.	N.S.	
	B Syn Dens	0.85	N.S.	Control	N.S.	N.S.	
	C T-unit	0.04	Sig.	Control	N.S.	Sig.	Con Girls
	D Fluency	0.39	N.S.	Exper.	N.S.	N.S.	

Questions	Null Hypotheses	Prob	Treatment Sig. at .05	Group Showing Greatest Gains	Teacher Interaction Significant at .05	Sex Interaction Significant at .05	Group Showing Greatest Gains
4	Gr 9 & 10 Mid-post						
	A Quality	0.96	N.S.	Control	N.S.	N.S.	
	B Syn Dens	0.18	N.S.	Exper.	N.S.	N.S.	
	C T-unit	0.39	N.S.	Exper.	N.S.	N.S.	
	D Fluency	0.04	Sig.	Control	N.S.	N.S.	
5	Gr 9 Mid-post						
	A Quality	0.74	N.S.	Control	N.S.	N.S.	
	B Syn Dens	0.96	N.S.	Control	N.S.	N.S.	
	C T-unit	0.56	N.S.	Control	N.S.	N.S.	
	D Fluency	0.40	N.S.	Control	N.S.	N.S.	
6	Gr 10 Mid-post						
	A Quality	0.69	N.S.	Exper.	N.S.	N.S.	
	B Syn Dens	0.10	N.S.	Exper.	N.S.	N.S.	
	C T-unit	0.15	N.S.	Exper.	N.S.	N.S.	
	D Fluency	0.03	Sig.	Control	N.S.	Sig.	Con Girls

Questions	Null Hypotheses	Prob	Treatment Sig. at .05	Group Showing Greatest Gains	Teacher Interaction Significant at .05	Sex Interaction Significant at .05	Group Showing Greatest Gains
7	Gr 9 & 10 Pre-post						
	A Quality	0.86	N.S.	Control	N.S.	N.S.	
	B Syn Dens	0.26	N.S.	Exper.	N.S.	N.S.	
	C T-unit	0.49	N.S.	Control	N.S.	Sig	Con Girls
	D Fluency	0.99	N.S.	Control	N.S.	N.S.	
8	Gr 9 Pre-post						
	A Quality	0.45	N.S.	Control	N.S.	N.S.	
	B Syn Dens	0.41	N.S.	Exper.	N.S.	N.S.	
	C T-unit	0.59	N.S.	Exper.	N.S.	N.S.	
	D Fluency	0.42	N.S.	Exper.	N.S.	N.S.	
9	Gr 10 Pre-post						
	A Quality	0.21	N.S.	Control	N.S.	Sig	Con Girls
	B Syn Dens	0.42	N.S.	Exper.	N.S.	N.S.	
	C T-unit	0.14	N.S.	Control	N.S.	Sig	Con Girls
	D Fluency	0.72	N.S.	Control	N.S.	Sig	Con Girls

revealed non-significant differences, 18 favored the control groups and 15 favored the experimental groups.

None of the 36 interactions between the treatment and the teacher who taught any one of the 4 pairs of experimental and control classes involved in the study was significant at the .05 level of confidence. However, 7 of the 36 interactions between the treatment and the sex of the subject were significant at the .05 level of confidence. All of these significant interactions were the result of the control girls' showing greater gains (or smaller losses) than the other 3 subgroups showed. These interactions can probably be attributed to regression towards the mean as the control girls scored lower on the pretest than the other 3 subgroups did.

APPENDIX C

Rerating of a sample of the compositions
rated in the quality analysis

Post-analysis discussion of the methods used to rate the compositions for the quality analysis and a re-examination of a sample of the compositions rated in the quality analysis suggested that the methods used might not have produced accurate ratings. Therefore, a sample of the compositions rated in the quality analysis was re-rated using two different methods of evaluation in an attempt to determine if more accurate ratings could be obtained.

For experimental purposes, a problem with the traditional method of rating compositions results from the order in which the compositions are rated. Since the pre- and posttests are randomly ordered before the rating sessions begin, raters compare a given student's composition with the compositions of the class as a whole rather than comparing the student's post-composition with his pre-composition. The raters assign a general value to each composition, a value which is both imprecise and colored by the preceding compositions. Consequently, if a student's posttest is the first composition rated and his pretest is the fortieth composition rated, the compositions might not be rated using the same criteria. It was thought, then, that a more accurate assessment of a student's growth in writing skill might be obtained by comparing his mid- and posttests with his pretest.[1]

---

[1]The investigator is grateful to Mrs. E.B.H. Martin for her insights into the rating problem, insights which led to the re-rating of the sample.

A sample of 54 compositions, the compositions rated on the first day of the original rating sessions was re-rated. The compositions of each student were placed in pre-mid-post order, the students' names were removed, and the sets of compositions written by experimental and control students were randomly intermingled. The raters were asked to compare the mid- and posttest each student wrote with his pretest and to evaluate the growth in writing skill demonstrated by each set of compositions on a scale of minus five to plus five. The raters felt that assigning numerical values to the compositions was not practical and consequently a second method was adopted. The raters were then asked to place the sets of compositions into one of two categories: sets of compositions showing growth in writing skill and sets of compositions showing little or no growth in writing skill. A third category was required for sets of compositions that the raters could not place in one of the other two categories.

If the treatment had influenced the writing of the experimental students, the compositions of these students should have shown growth in writing skill while the compositions written by the control students should have shown little or no growth in writing skill. As Table 141 shows, just over half of the nine sets of experimental compositions and just under half of the nine sets of control compositions showed growth in composition skill over the six months of the study, differences too small to be statistically significant. This finding merely confirmed the

results of the original analysis.

TABLE 141

Rerating of a sample of the compositions used in the quality analysis.

		Compositions Showing Growth		Compositions Showing No Growth		Undecided	
Rater	n	Exp	Con	Exp	Con	Exp	Con
1	18	5	3	4	4	0	2
2	18	5	3	4	4	0	2
Total	36	10	6	8	8	0	4

APPENDIX D

Revision of the Syntactic Density Score

Calculating the Syntactic Density Score: A Mathematical Problem[1]

Although the search for a reliable, quantitative measure of the maturity of written composition antedates LaBrant's (1933) subordination index, it was not until Hunt's (1965) identification of the T-unit that quantitative measures of sentence and clause length factors could be demonstrated to be valid, reliable indices of syntactic maturity.

Following Hunt's exploration of the T-unit, several researchers attempted to design measures embodying linguistic features beyond clause and sentence length factors which indicate the maturity of written composition. Such scales include Botel and Granowsky's (1972) "A formula for measuring syntactic complexity: A directional effort," Endicott's (1973) "A proposed scale for syntactic density" and Golub and Kidder's (1974) "syntactic density score."

The Golub and Kidder scale is recommended by its authors both for ease of scoring and for its empirical basis. The scale may be scored either by a computer program provided by the authors or by hand using a checklist (outlined in ERIC: ED 091 741). Roughly 1000 words per hour may be scored by hand. As explained by Golub and Kidder, the Syntactic Density Score (SDS) was designed by selecting 63 linguistic features from writing samples rated high, mid,

---

[1]J.F. Belanger, "Calculating the Syntactic Density Score: A Mathematical Problem," Research in the Teaching of English, Vol. 12 (May, 1978), pp. 149-153. Reprinted by permission of National Council of Teachers of English.

and low by teacher-raters. These 63 variables were subjected to multivariate analysis and the 10 salient features were then embodied in the SDS using loadings which reflect their strengths as predictors of good writing. In addition, the authors promise a vocabulary intensity scale to supplement their present syntactic measures.

Although the SDS appears to have a strong empirical base, researchers using the scale in its present form have reported several problems. O'Donnell (1976) suggested that there is a high degree of redundancy among the items measured, with a consequent high correlation between the SDS and T-unit measures. He also questioned the accuracy of the grade level conversion scale as a predictor of the grade level of a given composition.

However, a problem more immediate than these is a mathematical scoring anomaly in the scale whereby the SDS received by a writing sample depends more on the number of T-units analyzed than on any factor of syntactic density. Thus, if a researcher analyzes 5 T-units of a given passage he will arrive at a far higher score than if he had analyzed 10 T-units, which, in turn, is higher than if he had chosen 15 T-units.

The reason for this anomaly in scoring is that variables one to four (words per T-unit, subordinate clauses per T-unit, mean main clause word length, and mean subordinate clause word length) are 'averaged' scores and would be expected to remain the same whether 5, 10, or 15

T-units of a passage were analyzed. Variables five to ten (number of modals, be and have auxiliaries, prepositional phrases, possessives, adverbs of time, and gerunds, participles, and absolute phrases), on the other hand, are 'raw' scores and would be expected to vary with the length of the passage analyzed. Thus, if 16 T-units of a writing sample yield 18 prepositional phrases, 32 T-units of the same writing sample should yield 36 prepositional phrases.

To arrive at the SDS, the averaged values of variables one to four and the raw values of variables five to ten are summed and divided by the number of T-units analyzed. This averaging of weighted and unweighted numbers, a fallacy which mathematicians have no generic term to describe, results in an undependable quotient. The result in the SDS is that variables one to four, which remain constant independent of the number of T-units analyzed, create a false variation in the quotient as the denominator (the number of T-units analyzed) changes.

Table 142 illustrates the differing syntactic density scores that would be obtained by analyzing 8, 16, and 32 T-units from the same hypothetical passage. When all features of syntactic density are held constant, the shortest writing sample (Column V with 8 T-units) produces the highest SDS, a score almost twice as high as that of the 32 T-unit sample (Column VI).

As Table 143 illustrates, one approach to solving the mathematical anomaly in the SDS is to use raw scores rather

TABLE 142

Golub's Example

	I Raw Score	II Freq.	III Loading	IV F x L (Golub's Example)	V (Golub x .5)	VI (Golub x 2)	VII Weight in Percent (Col. IV)
Words				203	101.5	406	
T-units				16	8	32	
Variable*							
1.	203/16	12.7	.95	12.05	12.05	12.05	28.2
2.	14/16	.87	.90	.79	.79	.79	1.9
3.	131/16	8.2	.20	1.64	1.64	1.64	3.8
4.	88/14	6.3	.50	3.15	3.15	3.15	7.4
5.	5	5	.65	3.2	1.63	6.5	7.5
6.	5	5	.40	2.0	1.0	4.0	4.7
7.	18	18	.75	13.5	6.75	27.0	31.7
8.	4	4	.70	2.8	1.4	5.6	6.6
9.	3	3	.60	1.8	0.9	3.6	4.2
10.	2	2	.85	1.7	0.85	3.4	4.0
		Total		42.63	30.16	67.73	
Syntactic Density Score (Total/No. of T-units)				2.66	3.77	2.12	
Grade Level Conversion				4	5	3	

*Variables: 1. Words per T-unit, 2. Subordinate clauses per T-unit, 3. Main clause word length (mean), 4. Subordinate clause word length (mean), 5. Number of modals, 6. Number of <u>be</u> and <u>have</u> forms in the auxiliary position, 7. Number of prepositional phrases, 8. Number of possessives, 9. Number of adverbs of time, 10. Number of gerunds, participles, and absolute phrases (unbound modifiers).

Table 142: Columns I to IV reproduce data reported by Golub (1973) based on his analysis of a short writing sample. Column V, constructed hypothetically using Golub's data, assumes a writing sample half as long as Golub's (that is, eight T-units instead of 16) with exactly the same characteristics. Thus, while Golub counted 18 prepositional phrases in his sample (Column I, variable 7), the hypothetical writing sample, which is only eight T-units, includes nine prepositional phrases. Column VI assumes a writing sample twice as long as Golub's. Column VII reports the percentage that each variable contributes to the Syntactic Density Score (SDS) in Golub's example (Column IV). When all other factors are held constant, the SDS varies with the number of T-units analyzed.

TABLE 143

Raw Scores

	I (Golub's Example)	II (Golub x .5)	III (Golub x 2)	IV Weight in Percent
Words	203	101.5	406	
T-units	16	8	32	
Variable*				
1.	192.85	96.4	385.7	64.1
2.	12.6	6.3	25.2	4.2
3.	26.2	13.1	52.4	8.7
4.	44.0	22.0	88.0	14.6
5.	3.25	1.63	6.5	1.1
6.	2.0	1.0	4.0	0.7
7.	13.5	6.75	27.0	4.5
8.	2.8	1.4	5.6	0.9
9.	1.8	0.9	3.6	0.6
10.	1.7	0.85	3.4	0.6
Total	300.7	150.33	601.4	
SDS	18.79	18.79	18.79	

*Variable: Explained in Table 142.

Table 143: Columns I, II and III in Table 143 were constructed in a similar manner to Columns IV, V, and VI in Table 142. However, in this case variables one to four are adjusted raw score units: raw scores were multiplied by the loading factor but not divided by the number of T-units (variables one to three) or the number of subordinate clauses (variable four) as they were in Table 142. This modification assures that writing samples with similar characteristics but differing length receive similar scores. However, this modification unbalances the scale. T-unit length alone accounts for 64.1 percent of the score while other clause and sentence length factors account for an additional 27.5 percent.

than averaged scores for variables one to four. In this case all ten variables would have the same mathematical basis and the SDS received by a writing sample would not be dependent on the number of T-units analyzed. However, using raw scores for variables one to four would alter the percentage that each

variable contributes to the scale. Thus, while T-unit length accounts for 28.2 percent of the total score in Golub's original example (Table 142, Column IV), if raw scores were used, T-units would account for 64.1 percent of the scale (Table 143, Column 4). Furthermore, if raw scores were used, other clause and sentence length factors, variables two to four, would account for an additional 27.5 percent of the score compared to the 13.1 percent they contribute in Golub's example. Consequently, with clause and sentence length factors alone accounting for 91.6 percent of the total SDS, one might expect an almost perfect positive correlation between the SDS and T-unit length.

A more satisfactory method of solving the mathematical anomaly is to divide variables one to four by a constant.

As Table 144 indicates, dividing these variables by the constant 10 both removes the mathematical anomaly and restores the original balance of the scale. The 8, 16, and 32 T-unit samples from the same hypothetical writing sample all yield the same SDS, and, at the same time, the original balance of the scale, indicated by the profiles of percentage weightings in Columns IV and V, is restored.

Both O'Donnell's (1976) research and the example provided by Golub (1973) indicate that the grade level conversion table is invalid. O'Donnell reported that a sample of apparently average grade ten writing fell at the grade four level on the conversion table. Golub's example, containing sentences such as "We squatted down to look at Lemmus, slapping away at the mosquitoes which were eating our knees through our trousers, while Lemmus looked fiercely back

TABLE 144

Corrected Scores

	I (Golub's Example)	II (Golub x .5)	III (Golub x 2)	IV Weight in Percent	V Weight from Table 142
Words	203	101.5	406		
T-units	16	8	32		
Variable*					
1.	19.29	9.64	38.57	36.7	28.2
2.	1.26	0.63	2.52	2.4	1.9
3.	2.62	1.31	5.24	4.9	3.8
4.	4.4	2.2	8.8	8.4	7.4
5.	3.25	1.63	6.5	6.2	7.5
6.	2.0	1.0	4.0	3.8	4.7
7.	13.5	6.75	27.0	25.7	31.7
8.	2.8	1.4	5.6	5.3	6.6
9.	1.8	0.9	3.6	3.4	4.2
10.	1.7	0.85	3.4	3.2	4.0
Total	52.62	26.31	105.23		
SDS	3.29	3.29	3.29		

*Variable: Explained in Table 142.

Table 144: Table 144 was constructed in a similar manner to Table 143. However, in Table 144, variables one to four were divided by the constant 10. Dividing by any constant assures that length alone will not influence the SDS: writing samples with similar characteristics but different lengths will receive similar scores. Dividing by the constant 10 also restores the original balance of the scale as comparison of Columns IV and V indicates.

at us" also appears to be a considerably more mature passage than the grade four conversion it receives. At least part of this discrepancy, however, may be attributed to the number of T-units that the researchers analyzed. Had Golub chosen 5 T-units instead of 16, the conversion might have been more accurate.

Unfortunately, it is not clear that the above modifications to the SDS actually result in a valid measure of syntactic maturity since over 60-percent of the SDS is determined by two key factors, T-unit length and the number of prepositional phrases. Nevertheless, the modifications do remove the anomaly in scoring writing samples of different lengths while they retain the original balance of the scale, a balance which Golub and Kidder note was established through the analysis of a large number of writing samples.

REFERENCES

Botel, M., and A. Granowsky, A formula for measuring syntactic complexity: A directional effort, Elementary English, 1972, 49, 118-23.

Endicott, Anthony L., A Proposed scale for syntactic density, Research in the Teaching of English, 1973, 7, 5-12.

Golub, Lester S., Syntactic Density Score (SDS) with some aids for tabulating, 1973, ERIC: ED 091 741.

Golub, Lester S., and Carole Kidder, Syntactic density and the computer, Elementary English, 1974, 51, 1128-31.

Hunt, Kellogg W., Grammatical structures written at three grade levels, National Council of Teachers of English, Research Report No. 3, Champaign, Ill.: National Council of Teachers of English, 1965.

La Brant, Lou L. A study of certain language developments of children in grades 4-12 inclusive, Genetic Psychology Monographs, 1933, 14, 468-75.

O'Donnell, Roy C., A critique of some indices of syntactic maturity, Research in the Teaching of English, 1976, 10, 31-38.

APPENDIX E

Adjustment for topic difficulty:
the four major writing measures

Tables 145 and 146 give summaries of the average scores received by the grade-nine and grade-ten students respectively on Topics J (School), K (Home), and L (Society). Sample sizes, means, and standard deviations are given for each topic at each testing period for each of the four writing measures used in the study. The combined scores for the three testing periods are also given for each measure.

Tables 147 and 148 outline the methods used to calculate the topic adjustments (loading factors) for the grade-nine and grade-ten students respectively on each of the four writing measures. The topics with the highest average scores were selected on each measure and then the average scores on each of the other two topics were subtracted from this. The differences between the two topics was expressed as a percentage of the topic receiving the highest score. One point was added to this percentage to give the loading factor. Each student's score on each topic was multiplied by the loading factor for each measure prior to statistical analysis.

TABLE 145

Grade nine: average scores on each topic on each of the four writing measures

Grade 9	Topic J n	Topic J $\bar{x}$	Topic J σ	Topic K n	Topic K $\bar{x}$	Topic K σ	Topic L n	Topic L $\bar{x}$	Topic L σ
Quality									
first test	14	4.890	1.71	9	5.500	1.82	13	5.192	0.97
second test	12	5.125	1.09	12	6.083	1.36	12	5.417	0.85
third test	10	5.753	0.92	15	5.400	1.26	11	4.591	2.01
3 tests combined	36	5.208	1.34	36	5.652	1.44	36	5.083	1.35
Syntactic Density									
first test	17	3.444	0.62	15	3.343	0.71	16	3.769	0.70
second test	18	3.672	0.83	15	3.705	1.17	15	3.316	0.60
third test	13	3.667	0.90	18	3.447	0.89	17	3.238	0.79
3 tests combined	48	3.590	0.77	48	3.495	0.93	48	3.439	0.73
T-unit Length									
first test	37	14.461	2.36	34	13.208	2.58	32	14.821	3.11
second test	34	14.139	2.29	32	14.853	3.61	37	13.465	2.63
third test	32	14.511	2.71	37	14.982	3.40	34	14.121	2.68
3 tests combined	103	14.370	2.44	103	14.357	3.30	103	14.103	2.83
Fluency									
first test	37	252.243	102.24	34	267.265	83.70	32	232.188	72.94
second test	34	245.559	68.47	32	259.563	113.16	37	253.838	90.27
third test	32	239.094	83.35	37	195.351	60.84	34	220.706	88.72
3 tests combined	103	245.951	85.71	103	239.039	92.49	103	236.175	85.08

TABLE 146

Grade ten: average scores on each topic on each of the four writing measures

Grade 10	Topic J n	Topic J $\bar{x}$	Topic J σ	Topic K n	Topic K $\bar{x}$	Topic K σ	Topic L n	Topic L $\bar{x}$	Topic L σ
Quality									
first test	10	5.200	1.06	12	5.917	1.20	14	6.460	1.34
second test	15	5.833	1.99	10	6.450	1.34	11	5.455	2.32
third test	11	6.455	1.94	14	5.214	1.37	11	4.864	1.38
3 tests combined	36	5.847	1.78	36	5.792	1.37	36	5.667	1.79
Syntactic Density									
first test	14	3.599	1.46	16	3.731	0.91	18	3.503	0.86
second test	20	3.919	1.06	14	3.636	0.84	14	4.212	0.48
third test	14	3.982	0.59	18	3.425	0.78	16	3.966	1.02
3 tests combined	48	3.590	0.77	48	3.495	0.93	48	3.439	0.73
T-unit Length									
first test	33	14.984	3.56	29	15.425	2.98	29	14.529	2.89
second test	31	15.155	3.48	29	15.480	2.99	31	15.185	3.14
third test	27	15.634	3.61	34	13.569	2.38	30	15.090	3.76
3 tests combined	91	15.235	3.52	92	14.756	2.89	90	14.943	3.26
Fluency									
first test	33	274.000	105.74	29	252.793	103.37	29	258.414	102.51
second test	31	269.065	96.12	29	281.793	79.06	31	250.710	81.11
third test	27	273.926	100.71	34	219.971	103.80	30	238.000	82.21
3 tests combined	91	272.297	99.96	92	249.804	98.89	90	248.956	82.30

TABLE 147

Grade nine: Method of obtaining the loading factor for adjustment for topic difficulty

Measure	Average on topic with highest score	Average on other topics	Topic with highest score minus other topics	Difference as a percentage of topic with highest score	Loading factor (differences as a percentage plus one)
Quality	K (school) 5.6527	J (home) 5.2083	.4444	.085325	1.085325
		L (society) 5.0833	.5694	.112014	1.112014
Syntactic Density	J (home) 3.590	K (school) 3.4950	.0950	.027182	1.027182
		L (society) 3.439	.1510	.04391	1.04391
T-unit Length	J (home) 14.370	K (school) 14.357	.0130	.00091	1.00091
		L (society) 14.103	.2670	.01893	1.01893
Fluency	J (home) 245.951	K (school) 239.039	6.912	.02892	1.02892
		L (society) 236.175	9.776	.04139	1.04139

TABLE 148

Grade ten: Method of obtaining the loading factor for adjustment for topic difficulty

Measure	Average on topic with highest score	Average on other topics	Topic with highest score minus other topics	Difference as a percentage of topic with highest score	Loading factor (difference as a percentage plus one)
Quality	J (home) 5.847	K (school) 5.792	.055	.009496	1.009496
		L (society) 5.667	.180	.031763	1.031763
Syntactic Density	L (society) 3.864	J (home) 3.844	.020	.005202	1.005202
		K (school) 3.589	.275	.076623	1.076623
T-unit Length	J (home) 15.235	K (school) 14.756	.479	.032461	1.032461
		L (society) 14.943	.292	.019541	1.019541
Fluency	J (home) 272.297	K (school) 249.804	22.493	.090043	1.090043
		L (society) 248.956	23.341	.09376	1.09376

APPENDIX F

Adjustment for topic difficulty:
major errors

Tabulations of the numbers of errors on the three topics indicated that the three topics were not of equal difficulty. As Table 149 shows, the topic "school" elicited the highest number of errors on both the grade-nine and grade-ten levels while the topic "society" elicited the lowest number of errors. Consequently, the topics were adjusted by the loading factor given in Table 149 before the statistical computations were made.

The interesting observation here is that the topic "school" which appeared to be of middle difficulty on the other four writing measures elicited the most major errors per hundred words from both the grade-nine and grade-ten students in the study. On the other hand, the topic "society," the topic which appeared to be the most difficult topic on the other four writing measures, elicited the fewest major errors on both the grade-nine and grade-ten levels. One possible explanation of this is that the topic which demanded most of the student's concentration to discuss also received most of his attention to mechanics. On the other hand, it might also suggest that because the student used less complex sentence structure (as evidenced by scores on the T-unit measure, Tables 145 and 146) he had less opportunity for error.

TABLE 149

Method of obtaining the loading factor for adjustment for topic difficulty: major errors

Grade	Average on topic with highest score	Average on other topics	Topic with highest score minus other topics	Difference as a percentage of topic with highest score	Loading factor (difference as a percentage plus one)
10	K (school) 1.160835	J (home) 0.95658	0.204255	21.353	1.21353
		L (society) 0.917124	0.243711	26.573	1.26573
9	K (school) 0.965736	J (home) 0.918807	0.046929	5.10761	1.0510761
		L (society) 0.8430	0.122736	14.55944	1.1455944

APPENDIX G

Reading scores of the experimental and control groups by treatment, grade, and sex

TABLE 150

Reading scores of the boys and girls in each of the four experimental classes: average raw scores and change scores

Group	Sex	n	Test Averages: Pretest $\bar{x}$	Pretest σ	Midtest $\bar{x}$	Midtest σ	Posttest $\bar{x}$	Posttest σ	Change Scores: Pre-Mid	Mid-Post	Pre-Post
Teacher 1	M	10	47.10	11.18	58.10	10.74	52.10	9.48	11.00	-6.00	5.00
	F	13	52.69	16.84	55.69	15.17	52.77	19.08	3.00	-2.92	0.08
Teacher 2	M	15	49.20	13.92	52.07	14.32	51.87	14.01	2.87	-0.20	2.67
	F	6	48.17	21.91	54.33	16.78	51.17	20.51	6.17	-3.16	3.00
Grade 10	M	25	48.36	12.44	54.48	12.85	50.41	11.93	6.12	-4.07	2.05
	F	19	51.26	17.61	55.26	14.83	52.26	18.47	4.00	-3.00	1.00
Grade 10 Total		44	49.61	14.96	54.82	13.74	52.09	15.10	5.21	-2.73	2.48
Teacher 3	M	16	40.06	14.31	40.50	14.61	39.00	16.86	0.44	-1.50	-1.06
	F	13	45.31	14.15	42.85	15.27	48.69	15.01	-2.46	5.84	3.38
Teacher 4	M	12	20.75	9.86	21.08	9.67	24.08	11.93	0.33	3.00	3.33
	F	9	33.33	9.45	30.56	12.00	36.89	9.82	-2.76	6.32	3.56
Grade 9	M	28	31.79	15.47	32.18	15.60	32.61	16.21	0.39	0.43	0.82
	F	22	40.41	13.28	37.82	14.70	43.86	13.84	-2.59	6.04	3.45
Grade 9 Total		50	35.58	15.16	34.66	15.47	37.56	16.21	-0.92	2.90	1.98
Grades 9 & 10	M	53	39.60	16.36	42.70	18.18	41.74	17.30	3.10	-0.96	2.14
	F	41	45.44	16.36	45.90	17.13	47.76	16.69	0.46	1.86	2.32
Grades 9 & 10 Total		94	42.15	16.62	44.10	17.80	44.36	17.29	1.95	0.26	2.21

TABLE 151

Reading scores of the boys and girls in each of the four control classes: average raw scores and change scores

Group	Sex	n	Test Averages Pretest $\bar{x}$	Pretest σ	Midtest $\bar{x}$	Midtest σ	Posttest $\bar{x}$	Posttest σ	Change Scores Pre-Mid	Mid-Post	Pre-Post
Teacher 1	M	13	49.92	13.11	51.54	12.38	47.54	18.24	1.62	-4.00	-2.39
	F	10	50.70	18.39	51.40	16.97	48.00	17.49	0.70	-3.40	-2.70
Teacher 2	M	10	50.30	14.07	50.80	13.06	49.70	15.54	0.50	-1.10	-0.60
	F	14	58.64	13.88	57.93	19.22	60.14	15.00	-0.71	2.21	1.50
Grade 10	M	23	50.09	12.93	51.22	12.12	48.48	16.41	1.13	-2.74	-1.61
	F	24	55.33	15.70	55.21	17.84	55.08	16.51	-0.12	-0.13	-0.25
Grade 10 Total		47	52.77	14.65	53.26	15.44	51.85	16.79	0.49	-1.41	-0.92
Teacher 3	M	11	46.46	15.79	44.00	15.98	48.27	14.82	-2.46	4.27	1.82
	F	18	44.22	15.01	41.39	14.94	42.39	16.08	-2.83	1.00	-1.83
Teacher 4	M	13	42.39	14.24	41.69	14.06	41.08	12.02	-0.69	-0.62	-1.31
	F	11	36.91	13.08	34.55	13.84	37.18	16.62	-2.36	2.64	0.27
Grade 9	M	24	44.25	14.47	42.75	14.37	44.38	13.29	-1.50	1.63	0.13
	F	29	41.45	14.27	38.79	14.42	40.41	15.91	-2.66	1.62	-1.04
Grade 9 Total		53	42.72	14.43	40.59	14.53	42.21	14.91	-2.13	1.62	-0.51
Grades 9 & 10	M	47	47.11	14.05	46.89	13.97	46.38	15.04	-0.22	-0.51	-0.73
	F	53	47.74	16.46	46.23	18.02	47.06	17.75	-1.51	0.83	-0.68
Grades 9 & 10 Total		100	47.44	15.37	46.54	16.25	46.74	16.54	-0.90	0.20	-0.70

APPENDIX H

Rater agreement

TABLE 152

Agreement among the three raters on the quality scores. Grade-ten compositions were rated on days 1 and 2; grade-nine compositions were rated on days 3 and 4. Figures show the number of compositions out of 56 each day that raters assigned similar scores

	Compositions receiving two ratings		Compositions receiving three ratings			
Day	Raters A and B assigned the same score	Scores of A and B within one point	Rater C assigned the same score as rater A or B	Scores of rater C within one point of A or B	Agreement among three raters	Scores of raters A, B, and C more than one point apart*
1	18	16	11	7	52	2
2	13	27	4	7	51	3
3	16	19	7	10	52	2
4	21	18	5	10	54	0
Total	68/216	80/216	27/68	34/68	209/216	7/216

*The ratings for these seven cases are presented in Table 153.

TABLE 153

Scores assigned to the seven compositions on which the raters did not agree.

Student Number	Rater A	Rater B	Rater C	Group Mean*	Final Score**
15	4	8	6	6.2	5
25	8	4	6	6.2	5
63	6	10	8	5.5	9
82	7	5	2	5.5	4.5
103	2	4	8	5.5	5
140	5	0	2	5.4	1
144	8	6	4	5.4	6

*The mean was calculated for the compositions rated each day.

**The final score was computed using Diederich's (1974) method: the score of rater A or B closest to the group mean was discarded; the score of the other rater was averaged with the score of rater C.

APPENDIX I

Correlations between the reading and the four writing measures by grade, treatment, and sex

TABLE 154

Correlations between the reading and quality measures by grade, treatment, and sex for each of the three test periods.

Group	n	Pretest r	Pretest Prob	Midtest r	Midtest Prob	Posttest r	Posttest Prob
Total sample	72	.466	.0001	.535	.0001	.477	.0001
All boys	33	.352	.045	.326	.064	.372	.033
All girls	39	.571	.0001	.695	.0001	.587	.0001
All gr. 10	36	.494	.002	.651	.0001	.428	.009
Gr. 10 boys	19	.452	.052	.427	.068	.438	.061
Gr. 10 girls	17	.519	.033	.853	.0001	.540	.025
All Gr. 9	36	.400	.016	.443	.007	.582	.0001
Gr. 9 boys	14	.301	.295	.390	.168	.699	.005
Gr. 9 girls	22	.531	.011	.468	.028	.565	.006
All exp.	36	.362	.030	.225	.187	.295	.081
All con.	36	.564	.0001	.718	.0001	.603	.0001
Gr. 10 exp.	18	.482	.043	.282	.257	.315	.203
Gr. 10 con.	18	.491	.039	.850	.0001	.510	.031
Gr. 9 exp.	18	.164	.516	.293	.238	.341	.166
Gr. 9 con.	18	.600	.009	.556	.017	.744	.0001
All exp. boys	21	.391	.079	.222	.334	.397	.075
All exp. girls	15	.311	.260	.359	.188	.070	.804
All con. boys	12	.280	.378	.516	.086	.285	.370
All con. girls	24	.699	.0001	.792	.0001	.746	.0001
Gr. 10 exp. boy	13	.575	.040	.429	.143	.584	.036
Gr. 10 con. boy	6	.115	.828	.620	.189	.177	.738
Gr. 10 exp. girl	5	.516	.373	.653	.232	.132	.832
Gr. 10 con. girl	12	.673	.016	.934	.0001	.687	.014
Gr. 9 exp. boy	8	.489	.219	.550	.158	.552	.152
Gr. 9 con. boy	6	.402	.430	.354	.491	.744	.090
Gr. 9 exp. girl	10	.095	.794	.192	.596	.061	.868
Gr. 9 con. girl	12	.719	.008	.676	.016	.727	.007

TABLE 155

Correlations between the reading and syntactic density measures by grade, treatment, and sex for each of the three test periods.

Group	n	Pretest r	Pretest Prob	Midtest r	Midtest Prob	Posttest r	Posttest Prob
Total sample	96	.023	.823	.068	.512	.098	.341
All boys	45	.131	.391	.217	.153	.118	.441
All girls	51	-.064	.657	-.050	.726	.088	.541
All gr. 10	48	-.021	.886	-.017	.908	.138	.349
Gr. 10 boys	25	-.072	.732	.226	.277	.157	.452
Gr. 10 girls	23	.143	.514	-.201	.357	.115	.601
All gr. 9	48	.045	.763	.020	.893	.071	.634
Gr. 9 boys	20	.180	.449	.041	.863	-.021	.931
Gr. 9 girls	28	-.126	.524	.008	.968	-.071	.719
All exp.	48	.125	.398	.026	.859	.202	.169
All con.	48	-.095	.521	.084	.570	.017	.907
Gr. 10 exp.	24	.047	.826	-.039	.855	.138	.520
Gr. 10 con.	24	-.089	.679	-.016	.939	.139	.518
Gr. 9 exp.	24	.172	.422	.054	.801	.127	.555
Gr. 9 con.	24	-.155	.470	.004	.984	-.226	.289
All exp. boys	27	.273	.169	.096	.635	.234	.241
All exp. girls	21	-.091	.695	-.070	.762	.272	.233
All con. boys	18	-.113	.656	.382	.118	.068	.787
All con. girls	30	-.068	.720	-.059	.757	-.006	.974
Gr. 10 exp. boy	16	.099	.716	.132	.626	.138	.611
Gr. 10 con. boy	9	-.319	.403	.537	.136	.203	.600
Gr. 10 exp. girl	8	-.089	.833	-.441	.274	.138	.745
Gr. 10 con. girl	15	.226	.418	-.169	.548	.085	.764
Gr. 9 exp. boy	11	.254	.451	-.141	.680	.563	.071
Gr. 9 con. boy	9	-.013	.973	.193	.619	-.151	.699
Gr. 9 exp. girl	13	-.035	.910	.182	.553	.333	.226
Gr. 9 con. girl	15	-.208	.457	-.125	.658	-.259	.351

[illegible]

TABLE 156

Correlations between the reading and T-unit measures by grade, treatment, and sex for each of the three test periods.

Group	n	Pretest r	Pretest Prob	Midtest r	Midtest Prob	Posttest r	Posttest Prob
Total sample	194	.058	.426	.077	.283	.094	.193
All boys	100	.120	.236	.118	.240	.149	.140
All girls	94	-.009	.933	.041	.693	.038	.715
All gr. 10	91	.031	.770	.125	.239	.191	.070
Gr. 10 boys	48	.036	.806	.174	.237	.137	.352
Gr. 10 girls	43	.070	.654	.085	.589	.234	.132
All gr. 9	103	-.036	.722	-.170	.086	-.027	.783
Gr. 9 boys	52	-.011	.938	-.085	.549	.104	.461
Gr. 9 girls	51	-.013	.473	-.257	.069	-.161	.259
All exp.	94	.156	.134	-.024	.816	.135	.196
All con.	100	-.040	.690	.178	.077	.053	.599
Gr. 10 exp.	44	.078	.614	.091	.556	.140	.365
Gr. 10 con.	47	.013	.932	.163	.275	.233	.116
Gr. 9 exp.	50	.022	.879	-.308	.030	.079	.584
Gr. 9 con.	53	-.146	.297	-.012	.934	-.145	.300
All exp. boys	53	.145	.299	-.045	.751	.062	.660
All exp. girls	41	.131	.414	-.004	.981	.216	.175
All con. boys	47	.046	.759	.331	.023	.266	.071
All con. girls	53	-.110	.431	.077	.583	-.094	.505
Gr. 10 exp. boy	25	.005	.982	.015	.944	-.067	.751
Gr. 10 con. boy	23	.048	.827	.411	.051	.301	.163
Gr. 10 exp. girl	19	.121	.620	.168	.492	.285	.237
Gr. 10 con. girl	24	.143	.505	.035	.870	.202	.344
Gr. 9 exp. boy	28	.040	.841	-.257	.188	.190	.334
Gr. 9 con. boy	24	-.137	.523	.144	.502	.164	.444
Gr. 9 exp. girl	22	-.078	.729	-.363	.097	.052	.819
Gr. 9 con. girl	29	-.128	.509	-.152	.431	-.269	.158

TABLE 157

Correlations between the reading and fluency measures by grade, treatment, and sex for each of the three test periods.

Group	n	Pretest r	Pretest Prob	Midtest r	Midtest Prob	Posttest r	Posttest Prob
Total sample	194	.204	.004	.190	.008	.272	.0001
All boys	100	.255	.010	.152	.131	.359	.0001
All girls	94	.153	.142	.222	.032	.172	.097
All gr. 10	91	.221	.035	.331	.001	.236	.025
Gr. 10 boys	48	.232	.113	.119	.420	.216	.141
Gr. 10 girls	43	.243	.116	.451	.002	.277	.072
All gr. 9	103	.136	.172	-.004	.970	.215	.029
Gr. 9 boys	52	.195	.166	.038	.791	.342	.013
Gr. 9 girls	51	.057	.689	-.057	.693	.041	.773
All exp.	94	.193	.062	.340	.001	.334	.001
All con.	100	.205	.040	.066	.512	.208	.037
Gr. 10 exp.	44	.083	.594	.310	.040	.163	.292
Gr. 10 con.	47	.350	.016	.339	.020	.283	.054
Gr. 9 exp.	50	.184	.200	.240	.093	.345	.014
Gr. 9 con.	53	.025	.858	-.236	.088	.011	.939
All exp. boys	53	.134	.337	.332	.015	.416	.002
All exp. girls	41	.236	.137	.343	.028	.190	.233
All con. boys	47	.363	.012	-.061	.684	.273	.063
All con. girls	53	.096	.493	.143	.306	.160	.253
Gr. 10 exp. boy	25	.037	.862	.213	.307	.163	.435
Gr. 10 con. boy	23	.417	.048	-.003	.989	.277	.201
Gr. 10 exp. girl	19	.106	.665	.414	.078	.182	.456
Gr. 10 con. girl	24	.427	.038	.508	.011	.361	.083
Gr. 9 exp. boy	28	.135	.493	.332	.084	.428	.023
Gr. 9 con. boy	24	.235	.268	-.200	.348	.185	.388
Gr. 9 exp. girl	22	.238	.286	.147	.514	.154	.495
Gr. 9 con. girl	29	-.102	.600	-.273	.152	-.039	.841

APPENDIX J

Correlations among the writing measures
by grade and sex

TABLE 158

Correlations among the four writing measures for the boys in the study

Test		Quality			Syntactic Density			T-unit Length			Fluency		
		n	r	Prob	n	r	Prob	n	r	Prob	n	r	Prob
Quality	pre	33											
	mid	33											
	post	33											
Syntactic Density	pre	33	-.016	.928									
	mid	33	.014	.940									
	post	33	-.098	.589									
T-unit Length	pre	33	-.109	.546	45	.948	.0001						
	mid	33	.015	.934	45	.941	.0001						
	post	33	-.107	.554	45	.883	.0001						
Fluency	pre	33	.396	.023	45	-.108	.479	100	-.024	.813			
	mid	33	.390	.025	45	-.191	.209	100	-.186	.063			
	post	33	.463	.007	45	.155	.308	100	.116	.249			

TABLE 159

Correlations among the four writing measures for the girls in the study

Test		Quality			Syntactic Density			T-unit Length			Fluency		
		n	r	Prob	n	r	Prob	n	r	Prob	n	r	Prob
Quality	pre	39											
	mid	39											
	post	39											
Syntactic Density	pre	39	-.100	.545									
	mid	39	-.053	.747									
	post	39	.049	.769									
T-unit Length	pre	39	-.043	.797	51	.940	.0001						
	mid	39	-.012	.497	51	.873	.0001						
	post	39	.038	.819	51	.925	.0001						
Fluency	pre	39	.336	.036	51	-.023	.874	94	.046	.663			
	mid	39	.562	.0001	51	-.319	.023	94	-.164	.114			
	post	39	.382	.016	51	.030	.837	94	.069	.511			

TABLE 160

Correlations among the writing measures for the experimental groups in the study

Test		Quality			Syntactic Density			T-unit Length			Fluency		
		n	r	Prob	n	r	Prob	n	r	Prob	n	r	Prob
Quality	pre	36											
	mid	36											
	post	36											
Syntactic Density	pre	36	-.082	.636									
	mid	36	.109	.528									
	post	36	.106	.536									
T-unit Length	pre	36	-.101	.559	48	.935	.0001						
	mid	36	.046	.789	48	.850	.0001						
	post	36	.092	.594	48	.847	.0001						
Fluency	pre	36	.369	.027	48	-.280	.054	94	-.087	.405			
	mid	36	.448	.006	48	-.196	.183	94	-.159	.126			
	post	36	.594	.0001	48	-.051	.728	94	.029	.783			

TABLE 161

Correlations among the four writing measures for the control groups in the study

Test		Quality			Syntactic Density			T-unit Length			Fluency		
		n	r	Prob	n	r	Prob	n	r	Prob	n	r	Prob
Quality	pre	36											
	mid	36											
	post	36											
Syntactic Density	pre	36	-.065	.708									
	mid	36	-.106	.537									
	post	36	-.111	.519									
T-unit Length	pre	36	-.053	.760	48	.953	.0001						
	mid	36	-.144	.401	48	.940	.0001						
	post	36	-.120	.486	48	.944	.0001						
Fluency	pre	36	.337	.044	48	.054	.716	100	.106	.292			
	mid	36	.502	.002	48	-.297	.041	100	-.181	.072			
	post	36	.298	.078	48	.183	.213	100	.149	.140			

APPENDIX K

Means, Standard Deviations, and Change Scores on the Reading and Four Writing Measures

TABLE 162

Grade 10: Means and standard deviations on the reading and quality measures on the pre-, mid-, and posttests. Changes on the three test comparisons are presented below

Average Scores

	n	Pretest Writing $\bar{x}$	Pretest Writing σ	Pretest Reading $\bar{x}$	Pretest Reading σ	Midtest Writing $\bar{x}$	Midtest Writing σ	Midtest Reading $\bar{x}$	Midtest Reading σ	Posttest Writing $\bar{x}$	Posttest Writing σ	Posttest Reading $\bar{x}$	Posttest Reading σ
Experimental													
Boys	13	5.60	1.23	46.54	14.91	5.32	1.13	53.46	15.18	4.96	1.57	49.46	14.39
Girls	5	6.68	1.30	43.20	14.59	6.57	1.02	43.80	12.92	6.23	1.23	40.20	17.65
Combined	18	5.90	1.34	45.61	14.90	5.66	1.24	50.78	15.22	5.31	1.59	46.89	15.91
Control													
Boys	6	6.57	0.76	47.83	13.94	6.08	2.24	48.83	13.43	4.65	0.79	47.67	20.67
Girls	12	5.96	1.47	53.92	15.56	6.33	2.44	53.42	18.91	6.37	1.66	51.33	17.49
Combined	18	6.16	1.31	51.89	15.32	6.25	2.38	51.89	17.42	5.79	1.64	50.11	18.87

Change Scores

	n	Pre-Mid Comparison Writing	Pre-Mid Comparison Reading	Mid-Post Comparison Writing	Mid-Post Comparison Reading	Pre-Post Comparison Writing	Pre-Post Comparison Reading
Experimental							
Boys	13	-0.28	+6.92	-0.36	-4.00	-0.64	+2.92
Girls	5	-0.11	+0.60	-0.34	-3.60	-0.45	-3.00
Combined	18	-0.24	+5.17	-0.35	-3.89	-0.59	+1.28
Control							
Boys	6	-0.49	+1.00	-1.43	-1.16	-1.92	-0.16
Girls	12	+0.37	-0.50	+0.04	-2.09	+0.41	-2.59
Combined	18	+0.09	0.00	-0.45	-1.78	-0.37	-1.78

TABLE 163

Grade 10: Means and standard deviations on the reading and syntactic density measures of the pre-, mid-, and posttests. Changes on the three test comparisons are presented below

Average Scores

		Pretest				Midtest				Posttest			
		Writing		Reading		Writing		Reading		Writing		Reading	
	n	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
Experimental													
Boys	16	3.81	0.90	47.88	13.72	3.90	0.87	52.81	13.75	3.90	0.87	49.75	13.30
Girls	8	3.29	0.79	47.38	12.85	3.57	0.96	49.38	12.66	3.80	0.60	46.50	16.39
Combined	24	3.64	0.90	47.71	13.44	3.79	0.92	51.67	13.50	3.87	0.79	48.67	14.50
Control													
Boys	9	4.44	1.45	50.22	13.76	4.43	0.73	51.11	12.97	3.67	0.72	49.56	18.99
Girls	15	3.38	0.91	54.73	14.66	4.05	1.04	54.00	17.36	4.00	0.92	52.87	16.02
Combined	24	3.78	1.25	53.04	14.50	4.19	0.95	52.92	15.92	3.88	0.87	51.63	17.27

Change Scores

		Pre-Mid Comparison		Mid-Post Comparison		Pre-Post Comparison	
	n	Writing	Reading	Writing	Reading	Writing	Reading
Experimental							
Boys	16	+0.09	+4.93	+0.00	-3.06	+0.09	+1.87
Girls	8	+0.28	+2.00	+0.23	-2.88	+0.57	-0.88
Combined	24	+0.15	+3.96	+0.08	-3.00	+0.23	+0.96
Control							
Boys	9	-0.01	+0.89	-0.76	-1.55	-0.77	-0.66
Girls	15	+0.67	-0.73	-0.05	-1.13	+0.62	-1.86
Combined	24	+0.41	-0.12	-0.31	-1.29	+0.10	-1.41

TABLE 164

Grade 10: Means and standard deviations on the reading and T-unit measures on the pre-, mid-, and posttests. Changes on the three test comparisons are presented below

Average Scores

		Pretest				Midtest				Posttest			
		Writing		Reading		Writing		Reading		Writing		Reading	
	n	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
Experimental													
Boys	25	15.24	3.02	48.36	12.44	14.96	2.99	54.48	12.85	14.73	2.99	51.96	11.93
Girls	19	16.11	3.16	51.26	17.61	15.47	3.23	55.26	14.83	15.33	3.65	52.26	18.47
Combined	44	15.62	3.11	49.61	14.96	15.18	3.11	54.82	13.74	14.99	3.31	52.09	15.10
Control													
Boys	23	16.13	3.51	50.09	12.93	15.97	3.02	51.22	12.12	15.10	3.29	48.48	16.41
Girls	24	13.66	2.42	55.33	15.70	15.75	3.54	55.21	17.84	14.72	3.44	55.08	16.51
Combined	47	14.87	3.24	52.77	14.65	15.86	3.30	53.26	15.44	14.91	3.37	51.85	16.79

Change Scores

		Pre-Mid Comparison		Mid-Post Comparison		Pre-Post Comparison	
	n	Writing	Reading	Writing	Reading	Writing	Reading
Experimental							
Boys	25	-0.28	+6.12	-0.23	-2.52	-0.51	+3.60
Girls	19	-0.64	+4.00	-0.14	-3.00	-0.78	+1.00
Combined	44	-0.44	+5.21	-0.19	-2.73	-0.63	+2.48
Control							
Boys	23	-0.16	+1.13	-0.87	-2.74	-1.03	-1.61
Girls	24	+2.09	-0.12	-1.03	-0.13	+1.06	-0.25
Combined	47	+0.99	+0.49	-0.95	-1.41	+0.04	-0.92

TABLE 165

Grade 10: Means and standard deviations on the reading and fluency measures on the pre-, mid-, and posttests. Changes on the three test comparisons are presented below

Average Scores

		Pretest				Midtest				Posttest			
		Writing		Reading		Writing		Reading		Writing		Reading	
	n	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
Experimental													
Boys	25	264.7	108.4	48.36	12.44	286.7	72.4	54.48	12.85	255.4	93.5	51.96	11.93
Girls	19	300.5	90.5	51.26	17.61	318.4	79.3	55.26	14.83	257.9	72.8	52.26	18.47
Combined	44	280.1	102.6	49.61	14.96	300.4	77.0	54.82	13.74	256.5	85.2	52.09	15.10
Control													
Boys	23	305.9	109.2	50.09	12.93	290.4	58.38	51.22	12.12	274.4	107.3	48.48	16.41
Girls	24	244.5	108.8	55.33	15.70	244.1	121.9	55.21	17.84	240.2	119.2	55.08	16.51
Combined	47	274.6	113.2	52.77	14.65	266.8	98.9	53.26	15.44	256.9	114.8	51.85	16.79

Change Scores

		Pre-Mid Comparison		Mid-Post Comparison		Pre-Post Comparison	
	n	Writing	Reading	Writing	Reading	Writing	Reading
Experimental							
Boys	25	+22.0	+6.12	-31.3	-2.52	-9.30	+3.60
Girls	19	+17.9	+4.00	-60.5	-3.00	-42.6	+1.00
Combined	44	+20.3	+5.21	-43.9	-2.73	-23.6	+2.48
Control							
Boys	23	-15.5	+1.13	-16.0	-2.74	-31.5	-1.61
Girls	24	-0.4	-0.12	-3.9	-0.13	-4.3	-0.25
Combined	47	-7.8	+0.49	-9.9	-1.41	-17.7	-0.92

TABLE 166

Grade 9: Means and standard deviations on the reading and quality measures on the pre-, mid-, and posttests. Changes on the three test comparisons are presented below

Average Scores

		Pretest				Midtest				Posttest			
		Writing		Reading		Writing		Reading		Writing		Reading	
	n	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
Experimental													
Boys	8	6.16	1.32	33.38	11.77	5.95	0.90	35.25	11.26	5.35	1.70	29.13	13.29
Girls	10	5.22	0.72	39.90	12.72	5.74	1.23	38.50	13.43	5.48	0.80	42.10	12.81
Combined	18	5.64	1.13	37.00	12.73	5.83	1.10	37.06	12.62	5.42	1.28	36.33	14.53
Control													
Boys	6	5.64	2.15	46.67	12.26	5.77	1.04	44.33	12.71	6.79	1.65	46.50	11.80
Girls	12	5.30	1.72	41.42	13.43	6.04	1.25	37.67	15.53	5.10	1.60	37.41	16.57
Combined	18	5.41	1.88	43.17	13.28	5.95	1.19	39.89	14.98	5.64	1.80	40.45	15.74

Change Scores

		Pre-Mid Comparison		Mid-Post Comparison		Pre-Post Comparison	
	n	Writing	Reading	Writing	Reading	Writing	Reading
Experimental							
Boys	8	-0.21	+1.87	-0.60	-6.12	-0.81	-4.25
Girls	10	+0.52	-1.40	-0.26	+3.60	+0.26	+2.20
Combined	18	+0.19	+0.06	-0.41	-0.73	-0.22	-0.67
Control							
Boys	6	+0.13	-2.34	+1.02	+2.17	+1.15	-0.17
Girls	12	+0.74	-3.75	-0.94	-0.26	-0.20	-4.01
Combined	18	+0.54	-3.28	-0.31	+0.56	+0.23	-2.72

TABLE 167

Grade 9: Means and standard deviations on the reading and syntactic density measures on the pre-, mid-, and posttests. Changes on the three test comparisons are given below

Average Scores

		Pretest				Midtest				Posttest			
		Writing		Reading		Writing		Reading		Writing		Reading	
	n	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
Experimental													
Boys	11	3.34	0.61	28.82	12.90	3.73	0.67	30.73	13.44	3.85	0.69	25.55	13.18
Girls	13	3.64	0.69	42.69	14.37	3.71	0.83	39.31	15.34	3.27	0.68	43.54	14.39
Combined	24	3.51	0.67	36.33	15.36	3.72	0.76	35.38	15.12	3.54	0.74	35.29	16.50
Control													
Boys	9	3.45	0.45	44.56	12.93	3.69	1.17	43.44	13.53	3.34	1.07	44.00	10.59
Girls	15	3.86	0.83	44.67	14.81	3.65	0.86	40.27	16.39	3.60	0.86	39.93	18.22
Combined	24	3.70	0.74	44.63	14.13	3.67	0.99	41.46	15.45	3.50	0.96	41.46	15.92

Change Scores

		Pre-Mid Comparison		Mid-Post Comparison		Pre-Post Comparison	
	n	Writing	Reading	Writing	Reading	Writing	Reading
Experimental							
Boys	11	+0.39	+1.91	+0.12	-5.18	+0.51	-3.27
Girls	13	+0.07	-3.38	-0.44	+4.23	-0.37	+0.85
Combined	24	+0.21	-0.95	-0.18	-0.09	+0.03	-1.04
Control							
Boys	9	+0.24	-1.12	-0.35	+0.56	-0.11	-0.56
Girls	15	-0.21	-4.40	-0.05	-0.34	-0.26	-4.74
Combined	24	-0.03	-3.17	-0.17	0.00	-0.20	-3.17

TABLE 168

Grade 9: Means and standard deviations on the reading and T-unit measures on the pre-, mid-, and posttests. Changes on the three test comparisons are presented below

Average Scores

		Pretest				Midtest				Posttest			
		Writing		Reading		Writing		Reading		Writing		Reading	
	n	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
Experimental													
Boys	28	13.73	2.92	31.79	15.47	14.38	2.77	32.18	15.60	14.68	2.32	32.61	16.21
Girls	22	14.29	2.70	40.41	13.28	14.09	3.47	37.82	14.70	14.16	3.26	43.86	13.84
Combined	50	13.97	2.84	35.58	15.16	14.25	3.10	34.66	15.47	14.45	2.78	37.56	16.21
Control													
Boys	24	13.93	2.23	44.25	14.47	14.21	2.78	42.75	14.37	13.90	2.61	44.38	13.29
Girls	29	14.99	2.95	41.45	14.27	14.16	2.58	38.79	14.42	15.59	3.24	40.41	15.91
Combined	53	14.51	2.70	42.72	14.43	14.18	2.67	40.59	14.53	14.83	3.09	42.21	14.91

Change Scores

		Pre-Mid Comparison		Mid-Post Comparison		Pre-Post Comparison	
	n	Writing	Reading	Writing	Reading	Writing	Reading
Experimental							
Boys	28	+0.65	+0.39	+0.30	+0.43	+0.95	+0.82
Girls	22	-0.20	-2.59	+0.07	+6.04	-0.13	+3.45
Combined	50	+0.28	-0.92	+0.20	+2.90	+0.48	+1.98
Control							
Boys	24	+0.28	-1.50	-0.31	+1.63	-0.03	+0.13
Girls	29	-0.83	-2.66	+1.43	+1.62	+0.60	-1.04
Combined	53	-0.33	-2.13	+0.65	+1.62	+0.32	-0.51

TABLE 169

Grade 9: Means and standard deviations on the reading and fluency scores on the pre-, mid-, and posttests. Changes on the three test comparisons are presented below

Average Scores

		Pretest				Midtest				Posttest			
		Writing		Reading		Writing		Reading		Writing		Reading	
	n	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
Experimental													
Boys	28	241.3	91.4	31.79	15.47	258.4	80.8	32.18	15.60	198.2	82.6	32.61	16.21
Girls	22	250.7	102.3	40.41	13.28	265.3	106.5	37.82	14.70	232.2	96.2	43.86	13.84
Combined	50	245.4	96.5	35.58	15.16	261.4	93.1	34.66	15.47	213.1	90.4	37.56	16.21
Control													
Boys	24	256.7	71.6	44.25	14.47	249.9	110.2	42.75	14.37	208.9	65.6	44.38	13.29
Girls	29	275.4	86.0	41.45	14.27	262.2	77.7	38.79	14.42	249.3	65.5	40.41	15.91
Combined	53	266.9	80.3	42.72	14.43	256.6	94.0	40.59	14.53	231.0	68.5	42.21	14.91

Change Scores

		Pre-Mid Comparison		Mid-Post Comparison		Pre-Post Comparison	
	n	Writing	Reading	Writing	Reading	Writing	Reading
Experimental							
Boys	28	+17.1	+0.39	-60.2	+0.43	-43.1	+0.82
Girls	22	+14.6	-2.59	-33.1	+6.04	-18.5	+3.45
Combined	50	+16.0	-0.92	-48.3	+2.90	-32.3	+1.98
Control							
Boys	24	-6.8	-1.50	-41.0	+1.63	-47.8	+0.13
Girls	29	-13.2	-2.66	-12.9	+1.62	-26.1	-1.04
Combined	53	-10.3	-2.13	-25.6	+1.62	-35.9	-0.51

APPENDIX L

Rules for segmenting compositions into T-units

The following rules were used to segment the compositions into T-units:

1. An independent clause with all the subordinate clauses attached to it or embedded within it was counted as one T-unit.
2. Sentence fragments resulting from incorrect punctuation were attached (if possible) to a preceding or succeeding T-unit and counted as part of that T-unit. Fragments which could not be attached in this manner and fragments which resulted from missing words were treated as garbles (Hunt, 1965) and discarded.[1] The exception to this rule was the imperative sentence with "you" omitted: "come home by ten." This was counted as one four-word T-unit.
3. "So" was considered a coordinating conjunction unless the following clause strongly implied "so that" rather than "and so." The sentence "I always thought that school was made on a five day week so we could 'recouperate' (roughly

---

[1]O'Hare (1973) treated fragments which resulted from the omission of a word (or two? he does not say) as one T-unit. He supplied the missing word. Perhaps this was appropriate for the grade-seven sample that O'Hare was examining but it was not deemed appropriate for the grade-nine and -ten samples of the current study. Hunt (1965) discarded "garbles" but made no mention of sentences which could be corrected with the addition of one or two missing words. He noted, however, that investigators in his study agreed on T-unit segmentation "without exception so long as they were confronted by well-formed sentences. . ." (p. 22). This seems to imply that sentences which were not well-formed were discarded.

speaking) from the weeks work" was counted as one T-unit.

4. Following O'Hare (1973) "speaker tags" introducing or following direct discourse were counted as part of the direct discourse. "Where were you last night?' my father asked" was counted as one T-unit.

APPENDIX M

Initial equivalence of the experimental
and control groups

As noted in Chapter IV, the use of intact classes rather than fully-randomized groups required an assessment of the initial equivalence of the experimental and control groups on the reading and four writing measures. Comparison of the two groups by sex and by grade was done by means of t-tests.[1] The results of these analyses are presented below.

1. Reading. Comparison of the initial reading scores of the experimental and control groups indicates that these groups were not equal in reading ability at the beginning of the experiment. As Table 170 shows, on the initial test the control boys and the control girls on both the grade-nine and -ten levels scored higher than their experimental counterparts. However, only the grade-nine experimental boys, who scored over 12 points below the control boys, were significantly below their control counterparts. Although the initial reading scores of the experimental and control groups of grade-nine girls, grade-ten boys, and grade-ten girls were not significantly different, the substantial deficit of the grade-nine experimental boys

---

[1]Significance of the differences between the two groups was computed using Ferguson's formula for the significance of the difference between two means for independent samples (non-directional test), Ferguson (1971), pp. 169-71.

TABLE 170

Differences between the experimental and control groups on the initial reading measure

Group	Experimental $\bar{x}$	Experimental σ	Control $\bar{x}$	Control σ	Experimental minus control	t	df	decision	prob
Grade 9									
Boys	31.79	15.47	44.25	14.47	-12.46	-2.99	50	sig.	.01
Girls	40.41	13.28	41.45	14.27	-1.04	-0.27	49	n.s.	.80
Combined	35.58	15.16	42.72	14.43	-7.15	-2.45	101	sig.	.05
Grade 10									
Boys	48.36	12.44	50.09	12.93	-1.73	-0.47	46	n.s.	.70
Girls	51.26	17.61	55.33	15.70	-4.07	-0.79	41	n.s.	.50
Combined	49.61	14.96	52.77	14.65	-3.16	-1.02	89	n.s.	.40
Grades 9 & 10									
Boys	39.60	16.36	47.11	14.05	-7.51	-2.47	98	sig.	.05
Girls	45.44	16.36	47.74	16.46	-2.30	-0.67	92	n.s.	.50
Combined	42.15	16.62	47.44	15.37	-5.29	-2.30	194	sig.	.05

affected three of the combined groups: the grade-nine boys and girls combined, the combined grade-nine and grade-ten boys, and the combined grade-nine and grade-ten boys and girls. These three differences were significant beyond the .05 level of confidence.

2. Writing Quality. As Table 171 shows, there were no significant differences between the experimental and control groups on the initial writing quality measure. Unlike the reading scores noted above where all initial differences favored the control groups, two of the differences in writing quality, those for the grade-nine boys and the grade-ten girls, favored the experimental groups. The other two differences, the grade-nine girls and grade-ten boys, favored the control groups. Only one comparison, the grade-ten experimental and control boys, approached statistical significance, reaching the .07 level of confidence.

3. Syntactic Density. As Table 172 indicates, there were no significant differences between the experimental and control groups on the initial syntactic density measure. However, all control groups scored slightly higher than their experimental counterparts on this measure.

4. T-unit Length. As Table 173 indicates, the grade-

TABLE 171

Differences between the experimental and control groups on the initial writing quality measure

Group	Experimental		Control		Experimental				
	$\bar{x}$	σ	$\bar{x}$	σ	minus control	t	df	decision	prob
Grade 9									
Boys	6.16	1.32	5.64	2.15	.52	0.523	12	n.s.	.70
Girls	5.22	0.72	5.30	1.72	-.08	-0.146	20	n.s.	.90
Combined	5.64	1.13	5.41	1.88	.23	0.455	34	n.s.	.70
Grade 10									
Boys	5.60	1.23	6.57	0.76	-.97	-2.100	16	n.s.	.07
Girls	6.68	1.30	5.96	1.47	.72	1.000	15	n.s.	.40
Combined	5.90	1.34	6.16	1.31	-.26	-0.589	34	n.s.	.60
Grades 9 & 10									
Boys	5.81	1.29	6.10	1.68	-.29	-0.517	31	n.s.	.70
Girls	5.71	1.17	5.63	1.63	.08	0.178	37	n.s.	.90
Combined	5.77	1.24	5.79	1.66	-.02	-0.067	70	n.s.	.95

TABLE 172

Differences between the experimental and control groups on the initial syntactic density measure

Group	Experimental $\bar{x}$	Experimental σ	Control $\bar{x}$	Control σ	Experimental minus control	t	df	decision	prob
Grade 9									
Boys	3.34	.61	3.45	0.48	-.11	-0.451	18	n.s.	.70
Girls	3.65	.69	3.86	0.83	-.21	-0.731	26	n.s.	.50
Combined	3.51	.67	3.70	0.74	-.19	-0.932	46	n.s.	.40
Grade 10									
Boys	3.81	.90	4.44	1.45	-.63	-1.182	23	n.s.	.30
Girls	3.29	.79	3.38	0.91	-.09	-0.246	21	n.s.	.90
Combined	3.64	.90	3.78	1.25	-.14	-0.445	46	n.s.	.70
Grades 9 & 10									
Boys	3.62	.83	3.94	1.18	-.32	-0.997	43	n.s.	.40
Girls	3.51	.75	3.62	0.90	-.11	-0.474	49	n.s.	.70
Combined	3.57	.80	3.74	1.03	-.17	-0.903	94	n.s.	.40

TABLE 173

Differences between the experimental and control groups on the initial T-unit measure

Group	Experimental $\bar{x}$	Experimental σ	Control $\bar{x}$	Control σ	Experimental minus control	t	df	decision	prob
Grade 9									
Boys	13.73	2.92	13.93	2.23	-0.20	-0.280	50	n.s.	.80
Girls	14.29	2.70	15.00	2.95	-0.71	-0.893	49	n.s.	.40
Combined	13.97	2.84	14.51	2.70	-0.54	-0.998	101	n.s.	.40
Grade 10									
Boys	15.24	3.02	16.13	3.51	-0.89	-0.938	46	n.s.	.40
Girls	16.11	3.16	13.66	2.42	2.45	2.793	41	sig.	.01
Combined	15.62	3.11	14.87	3.24	0.75	1.127	89	n.s.	.30
Grades 9 & 10									
Boys	14.44	3.06	15.00	3.12	-0.56	-0.904	98	n.s.	.40
Girls	15.13	3.06	14.39	2.80	0.74	1.206	92	n.s.	.30
Combined	14.74	3.08	14.68	2.97	0.06	0.138	192	n.s.	.90

ten experimental and control girls were the only two apparently comparable groups which had, in fact, significantly different scores on the T-unit length measure on the initial test. These differences favored the experimental groups and were significant beyond the .01 level of confidence. The control students in the other three subgroups (the grade-nine boys, the grade-nine girls, and the grade-ten boys) scored higher than the experimental students although none of these differences approached statistical significance. When the grade-nine and grade-ten groups were combined, the experimental girls scored higher than their control counterparts while the experimental boys scored lower than the control boys did, but neither of these differences was statistically significant.

5. Fluency. As Table 174 shows, no statistically significant differences were observed between the experimental and control groups on the initial fluency measure although the control groups generally wrote longer compositions than their experimental counterparts wrote. The grade-ten experimental girls were the only experimental sub-group to score higher on the initial fluency measure than their control counterparts did. The other subgroups, the grade-nine experimental boys,

TABLE 174

Differences between the experimental and control groups on the initial fluency measure

Group	Experimental $\bar{x}$	Experimental σ	Control $\bar{x}$	Control σ	Experimental minus control	t	df	decision	prob
Grade 9									
Boys	241.26	91.43	256.73	71.61	-15.47	-0.684	50	n.s.	.50
Girls	250.71	102.27	275.37	85.98	-24.66	-0.913	49	n.s.	.40
Combined	245.42	96.46	266.93	80.33	-21.51	-1.230	101	n.s.	.30
Grade 10									
Boys	264.71	108.42	305.91	109.24	-41.20	-1.310	46	n.s.	.20
Girls	300.45	90.52	244.54	108.79	55.91	1.839	41	n.s.	.10
Combined	280.14	102.61	274.57	113.24	5.57	0.246	89	n.s.	.90
Grades 9 & 10									
Boys	252.32	100.49	280.79	95.20	-28.47	-1.454	98	n.s.	.20
Girls	273.76	100.12	264.41	98.18	12.35	0.598	92	n.s.	.60
Combined	261.67	100.89	270.52	97.27	-8.85	-0.621	192	n.s.	.60

the grade-nine experimental girls, and the grade-ten experimental boys, scored below their control counterparts on this measure, but not significantly.

APPENDIX N

An Examination of the Samples Used for the Quality and Syntactic Density Analyses

As was discussed in Chapter III, a 37-percent sample of the compositions gathered in the study was analyzed for overall writing quality while a 50-percent sample was analyzed for syntactic density. Since it was possible that these samples were not representative of the experimental and control groups as a whole, the students included in the sample groups were compared with the remainder of the students in the experiment to give an estimate of the representativeness of the samples. t-tests[1] were used to examine the statistical significance of the differences between the sample groups and the remainder of the population on the initial reading test, on the three-month reading gains scores and on the six-month reading gains scores. Comparisons were made for the experimental and control groups, for the grade-nine and grade-ten students separately, and for the grade-nine and grade-ten groups combined.

1. Quality Sample: Initial Reading Scores. As Table 175 indicates, there were no significant differences between the sample chosen for the quality analysis and the remainder of the students in the study when the initial reading scores of the two groups are compared. There were no

---

[1]Significance of the differences between the two groups was computed using Ferguson's formula for the significance of the difference between two means for independent samples (non-directional test), Ferguson (1971), pp. 169-71.

TABLE 175

Differences between the samples chosen for the quality analysis and the remainder of the experimental population on the initial reading measure

Group	Random sample			Remainder of population			Sample minus				
	n	$\bar{x}$	σ	n	$\bar{x}$	σ	remainder	t	df	decision	prob
Experimental											
Grade 9	18	37.00	12.73	32	34.78	16.58	2.22	0.497	48	n.s.	.70
Grade 10	18	45.61	14.90	26	52.39	14.65	-6.78	-1.478	42	n.s.	.20
Combined	36	41.31	14.51	58	42.67	17.93	-1.36	-0.386	92	n.s.	.70
Control											
Grade 9	18	43.17	13.28	35	42.49	15.20	0.68	0.163	51	n.s.	.90
Grade 10	18	51.89	15.32	29	53.31	13.56	-1.42	-0.323	45	n.s.	.80
Combined	36	47.53	14.98	64	47.39	15.36	0.14	0.044	98	n.s.	.99

TABLE 176

Differences between the samples chosen for the syntactic density analysis and the remainder of the experimental population on the initial reading measure

Group	Random sample			Remainder of population			Sample minus				
	n	$\bar{x}$	σ	n	$\bar{x}$	σ	remainder	t	df	decision	prob
Experimental											
Grade 9	24	36.33	15.36	26	34.89	15.24	1.44	.336	48	n.s.	.80
Grade 10	24	47.71	13.44	20	51.90	16.73	-4.19	-.925	42	n.s.	.40
Combined	48	42.02	15.51	46	42.28	17.89	-0.26	-.076	92	n.s.	.99
Control											
Grade 9	24	44.63	14.13	29	41.14	14.74	3.49	.876	51	n.s.	.40
Grade 10	24	53.04	14.50	23	52.39	14.19	0.65	.152	45	n.s.	.90
Combined	48	48.83	14.92	52	46.12	15.43	2.71	.881	98	n.s.	.40

significant differences on the grade-nine or grade-ten level for either the experimental or the control groups. Although the grade-ten experimental sample scored considerably below the remainder of the grade-ten experimental students in the study, this difference did not approach statistical significance. Since the sample for the grade-ten control group was also below the remainder of the grade-ten control group, these differences would tend to offset each other.

2. Syntactic Density Sample: Initial Reading Scores. As Table 176 indicates, there were no significant differences between the sample chosen for the syntactic density analysis and the remainder of the students in the study when the initial reading scores of the two groups were compared. There were no significant differences on either grade level for either the experimental or the control groups. Similar to the results reported for the quality sample above, the grade-ten experimental sample scored considerably below the remainder of the grade-ten experimental students in the study but this difference did not approach statistical significance.

3. Quality Sample: Three-Month Reading Gains Scores. As Table 177 indicates, the sample taken for the quality analysis was representative of the

TABLE 177

Differences between the samples chosen for the quality analysis
and the remainder of the experimental population
in pre-mid reading gains

Group	Random sample			Remainder of population			Sample minus				
	n	$\bar{x}$	σ	n	$\bar{x}$	σ	remainder	t	df	decision	prob
Experimental											
Grade 9	18	0.06	5.57	32	-1.57	7.34	1.63	.841	48	n.s.	.50
Grade 10	18	5.17	9.49	26	5.23	7.65	-0.06	-.024	42	n.s.	.99
Combined	36	2.61	8.10	58	1.59	8.17	1.02	.599	92	n.s.	.60
Control											
Grade 9	18	-3.28	5.51	35	-1.57	8.13	-1.71	-.804	51	n.s.	.50
Grade 10	18	0.00	8.62	29	0.79	6.89	-0.79	-.350	45	n.s.	.80
Combined	36	-1.64	7.32	64	-0.50	7.63	-1.14	-.730	98	n.s.	.50

groups as a whole when the reading gains scores for the first three months of the experiment were compared. At the .05 level of confidence there were no statistically significant differences between the samples chosen and the remainder of the students in the study when the grade-nine and grade-ten subjects were compared separately or when the two grades were combined.

4. <u>Syntactic Density Sample</u>: <u>Three-Month Reading Gains Scores</u>. As Table 178 indicates, the sample taken for the syntactic density analysis was representative of the groups as a whole when the reading gains scores for the first three months of the experiment were compared. At the .05 level of confidence there were no statistically significant differences between the samples chosen and the remainder of the students in the study when the grade-nine and grade-ten subjects were examined separately or when the two grades are combined.

5. <u>Quality Sample</u>: <u>Six-Month Reading Gains Scores</u>. As Table 179 shows, there were no statistically significant differences between the sample chosen for the quality analysis and the remainder of the students in the study when the six-month reading gains scores of the two groups are compared. In this case all of the sample groups showed smaller

TABLE 178

Differences between the samples chosen for the syntactic density analysis and the remainder of the experimental population in pre-mid reading gains

Group	Random Sample			Remainder of Population			Sample minus				
	n	$\bar{x}$	σ	n	$\bar{x}$	σ	remainder	t	df	decision	prob
Experimental											
Grade 9	24	-0.96	6.71	26	-0.96	6.86	0.00	0.000	48	n.s.	.99
Grade 10	24	3.95	8.74	20	6.70	7.81	-2.75	-1.090	42	n.s.	.30
Combined	48	1.50	8.10	46	2.52	8.19	-1.02	-0.616	92	n.s.	.60
Control											
Grade 9	24	-3.17	5.72	29	-1.31	8.45	-1.86	-0.920	51	n.s.	.40
Grade 10	24	-0.13	7.85	23	1.13	7.27	-1.26	-0.572	45	n.s.	.60
Combined	48	-1.65	6.97	52	-0.23	7.97	-1.42	-0.946	98	n.s.	.40

TABLE 179

Differences between the samples chosen for the quality analysis and the remainder of the experimental population in pre-post reading gains

Group	Random Sample n	Random Sample $\bar{x}$	Random Sample σ	Remainder of Population n	Remainder of Population $\bar{x}$	Remainder of Population σ	Sample minus remainder	t	df	decision	prob
Experimental											
Grade 9	18	-0.67	7.48	32	3.47	8.10	-4.14	-1.740	48	n.s.	.10
Grade 10	18	1.22	6.88	26	3.38	7.60	-2.16	-0.964	42	n.s.	.40
Combined	36	0.28	7.15	58	3.43	7.81	-3.15	-1.930	92	n.s.	.10
Control											
Grade 9	18	-2.72	7.86	35	0.57	8.60	-3.29	-1.347	51	n.s.	.20
Grade 10	18	-1.83	9.02	29	-0.03	7.64	-1.80	-0.731	45	n.s.	.50
Combined	36	-2.28	8.35	64	0.30	8.12	-2.58	-1.490	98	n.s.	.20

reading gains scores than the remainder of the students. Although the differences for the grade-nine experimental and control groups were quite large, both the experimental and control groups showed lower scores than the remainder of the students and these differences can be expected to offset each other.

6. Syntactic Density Sample: Six-Month Reading Gains Scores. As Table 180 shows, the sample populations chosen for the syntactic density analysis were not representative of the classes as a whole when the six-month reading gains scores of the two groups were compared. Three sample groups, the grade-nine experimental groups, the grade-nine control groups, and the grade-nine and -ten combined experimental groups, were significantly different from the remainder of the students in the study at or beyond the .05 level of confidence. However, since these differences were in the same direction for the experimental and control students these differences were not expected to bias the results of the study.

TABLE 180

Differences between the samples chosen for the syntactic density analysis and the remainder of the experimental population in pre-post reading gains

Group	Random Sample			Remainder of Population			Sample minus				
	n	$\bar{x}$	σ	n	$\bar{x}$	σ	remainder	t	df	decision	prob
Experimental											
Grade 9	24	-1.04	6.79	26	4.77	8.25	-5.81	-2.547	48	sig.	.03
Grade 10	24	0.92	7.10	20	4.40	7.28	-3.48	-1.570	42	n.s.	.20
Combined	48	-0.06	6.94	46	4.61	7.76	-4.67	-2.947	92	sig.	.01
Control											
Grade 9	24	-3.17	9.21	29	1.62	7.18	-4.79	-2.060	51	sig.	.05
Grade 10	24	-1.46	8.36	23	0.04	8.03	-1.50	-0.626	45	n.s.	.60
Combined	48	-2.31	8.74	52	0.92	7.53	-3.23	-1.947	98	n.s.	.10

APPENDIX O

Instructions to Students and Syntactic Density Score Coding Sheets

INSTRUCTIONS: STUDENT COMPOSITIONS

A. General:

Please put the following information at the top of each page:

1. your name
2. your class
3. your school

You will have 32 minutes to write an essay on the topic given below. I will tell you when three minutes remain so that you will be able to finish your essay within this time limit.

Because of the short time available, I recommend that you do not take time to look up words in the dictionary. Instead, if there is a word whose spelling you are not certain of and which you would normally look up in the dictionary, please underline it twice.

Your paper does not need to be any particular length. Write as much as you have to say about the topic.

Please do your best. Your essays may help us to find ways to make writing easier for students.

If you would like to make any notes before writing, please use the bottom (or back) of this sheet for your notes.

B. Specific:

Please write your essay on the topic "The Perfect (or Ideal) Country or Society." You may choose to write about your own country or how you would make this the ideal

country in which to live, or you may prefer to write about an imaginary country where you would like to live. You may write about this country in any way you wish. You may describe this country or tell what it would be like to live there. You may include both real and imaginary details.

SYNTACTIC DENSITY CODING SHEET

NAME: ____________________

#: ____________________

TOTAL WORDS: __________

TOTAL T-UNITS: ________

	freq	load	fX1
1. Wds/T-Unit	____	.95	____
2. SubCl/T-Unit	____	.90	____
3. MainCl Word Length (mean)	____	.20	____
4. SubCl Word Length (mean)	____	.50	____
5. # Modals	____	.65	____
6. # BE and HAVE in aux position	____	.40	____
7. # Prep Phrases	____	.75	____
8. # Possessives	____	.70	____
9. # Adv of Time	____	.60	____
10.# Gerunds, Participles, Absolute Phrases	____	.85	____

Total ________

SDS ________

APPENDIX P

Typical Good, Average and Poor Compositions

A. Grade-Nine Compositions

1. Good Composition: Experimental Boy. (Pretest) #161.

The Perfect Home

The perfect home is one that is trouble free. All the things that have to be done except for minor things are done by a machine. The minor things are things such a combing your hair, washing your face, brushing your teeth and walking. To wake up a claw will come from your wall a tickle you to get up. As this is happening a radio has been turned on. When you get up you have to walk into the bathroom and wash your face and comb your hair. While you are doing this a machine is making your bed and your radio is turned off automatically. When you are ready to have breakfast you walk into the kitchen and press a button, there is about 3 buttons to push one says "Bacon and eggs with milk", the other says "Toast, coffee and orange juice. The last one says "Poridge with orange juice and toast." After you eat you go to your bedroom and get dressed. When you reach your room your clothes are already set out for you, done by a machine that has been fed the information onto tapes that tell the machine what you like to wear. If the machine sets out the wrong clothes you just have to get them yourself. The house has an indoor swimming pool, a sun roof, a sauna and a sunken bathtub for your enjoyment. The house is equiped with a gymnastics room and a weight lifting room. In the winter you don't have to worry about shoveling snow because the sidewalks are

heated by an electrical current running underneath them. In the summer the grass usually gets long but the type of grass on this lawn stays at one length. This would be the perfect house. (Score 8.5).

2. Average Composition: Experimental Boy. (Midtest) #116.

The Ideal School

The ideal school I had in mind would have to be very special. First of all the teachers would have to be trained in a little bit of consuling work to asure the student she/or he would be able to discuss personel problems. Second would be attendance. The student would be allowed 40 days out of the school year to miss. Only allowed in high school. After the 40 days are up he would have to start bringing notes. Notes are a nusinsance to begin with. Thirdly, and I feel this very important the teachers discuss what they are allowed to take this year, the student's think up projects to do with each of these things. Long projects are easier to work on then "one-a-day" assignments. The fourth thing is to teach sex education, I think it is essential for the student body to know and learn more about sex. The best place to learn would be in school. The school is a place of learning and education, but that can turn out to be a damn drag sometimes. Giving longer breaks, say 5 minutes between each class instead of two would begin to make school easier. Better facilities would be nicer, such as a better library,

drama room, music room and the gym and bathrooms some of these places just dosen't suit the student's mind for creating. The ideal school would be the perfect thing that would ever happen to the student body. With clubs and the suggestions I have made in this essay, would be the ideal school for any student body. (Score 5).

3. <u>Poor Composition</u>: <u>Control Girl</u>. (Posttest) #140.

<u>The Perfect Country or Society</u>

It's fine the way it is. We don't need anybody trying to change the country. If some one made a perfect society in their point of view: no theifes, killers, crazy people, no cars trains, population control etc. life would be boring, nothing to watch or talk about. We would end up following schedules on what to do each day to the exact minute and I hate to follow schedules. If I was to write what I think a perfect society is, a teacher or parent or some tattle tail kid would read it and say some thing and I would be kicked out of school. So I won't write what I think would be a Perfect Society. Adults are not kids and never will be. Very few adults would understand why I would like to have certain things in a Perfect Society. If I did write what I had to say I may not be kicked out but teachers would treat me so bad I would have to quit. You said that this essay was to help find a way to make writing easier for students. Why don't anybody ever have <u>confedental</u> seacret essays. You know the kind that only one teacher reads and doen't gossip

about it with other teachers. Kids would write so much more in what they belive and know about things. And they won't have to worry about walking down the halls getting funny looks from teachers. The teacher that is marking the essay sould be ready to read anything and don't have a heart attack if a kid writes something against the law or swears. All I have to say is that teachers should be more conifidental and be preparede to read anything. (Score 1).

B. Grade-Ten Compositions

1. Good Composition: Control Girl. (Midtest) #11.

The Perfect Family

There is no such thing as the perfect family but this is my impression of the best family that anyone could find.

The best kind of family to live in is one where there is consideration for others. If one person wants to study for a test, the other members of the family should be willing to allow that person time to study by being quiet.

Everyone should be able to kid each other without anyone being offended. Joking about themselves ties a family closer together.

Love is essential. Parents must not show favortism towards one child over the other. The whole family must love and care for each other in every situation that may arise.

Living in a good family means sacrificing some things.

You may have to sacrifice your favorite television show to help paint the laundry room or you may have to sacrifice an hour of listening to your records because your mother has a horrible headache. If everyone is willing to make a few sacrifices in order to keep family harmony, then family ties will be closer.

A sign of frustration and a lack of family togetherness is often indicated by various members of the family physically combating one another. When fights occur, there should be a family conference so that the problem or problems of each member involved in the fight can be brought out into the open.

In my opinion, the best family is one that will stick together and stick up for each other in times of crisis. A crisis will prove the difference between a good family and a family that is not close together.

In conclusion, as I mentioned previously, there is no perfect family. But, if at least most of the examples of characteristics of good families that I have given are true of your family, then you have as near a perfect family as anyone can have. (Score 9).

2. <u>Average Composition</u>: <u>Experimental Boy</u>. (Midtest) #90.

<u>The Ideal School</u>

An ideal school would have to be one that would give the student a change to get involved with his work.

Most schools now have strict <u>coriculums</u> where a student must take this or this, and obviously a student who

hates and was not good in one subject will probably do very poor in it, thus lowering his chance for a job. Whereas if a student was able to choose what he feels he might need in order to get a job he would probably chanel his efforts to those subjects thus giving him some of the training he might need in order to continue that feild. For example a kid wanting to take welding as a job would now be forced to take social, but when is a weilder going to use social on a job, I' am sure that his employer is not going to ask him where Luxemburg is or whats this river called, so why should he be forced to take it.

The schools are gradully changing though, but still are not accepting the fact that not everybody going to go university or college. (Score 5).

3. Poor Composition: Experimental Boy. (Posttest) #98.

The Perfect Country

The perfect country would most likely be a free one. Free of buildings, streets and litter. It would be smoth, hilly, rough and green. The country would be rich in soil so the farmmers would be pleased year after year. It would also have room to roam around on with bikes. Hiking trails all over with lots of streams of fresh water. In the winter there would be lots of ski trails. (Score 2).